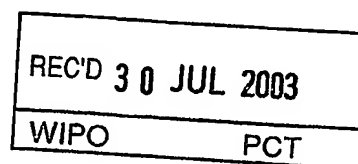


PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH
RULE 17.1(a) OR (b)



Kongeriget Danmark

Patent application No.: PA 2002 01500

Date of filing: 07 October 2002

Applicant:
(Name and address) Aarhus Universitet
Nordre Ringgade 1
DK-8000 Aarhus C
Denmark

Arbejdsmiljøinstituttet
Lersø Parkallé 105
DK-2100 København Ø
Denmark

Title: Cancer risk estimating method using sequence polymorphisms in a specific region of chromosome 19

IPC: -

This is to certify that the attached documents are exact copies of the above mentioned patent application as originally filed.

Patent- og Varemærkestyrelsen
Økonomi- og Erhvervsministeriet

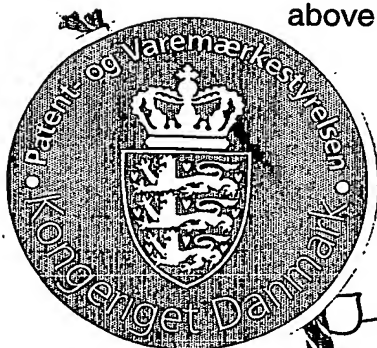
23 July 2003

Helle Schackinger Olesen

Helle Schackinger Olesen



PATENT- OG VAREMÆRKESTYRELSEN
BEST AVAILABLE COPY



P 687 DK01

- 7 OKT. 2002

1

PVS

Cancer risk estimating method using sequence polymorphisms in a specific region of chromosome 19

5 The present invention provides methods and compositions for identifying human subjects with an increased risk of having or developing cancer. In particular, this invention relates to the identification and characterization of polymorphisms in the human chromosome 19q, the region r located approximately 19q13.2-3 correlated with increased risk of developing cancer and the responsiveness of a subject to various treatments for cancer.

10

Background

15 DNA polymorphisms provide an efficient way to study the association of genes and diseases by analysis of linkage and linkage disequilibrium. With the sequencing of the human genome a myriad of hitherto unknown genetic polymorphisms among people have been detected. Most common among these are the single nucleotide polymorphisms, also called SNPs, of which we now know several millions. Other examples are variable number of tandem repeat polymorphisms, insertions, deletions and block modifications. Tandem repeats often have multiple different alleles (variants), whereas the other groups of polymorphisms usually just have two alleles. 20 Some of these genetic polymorphisms probably play a direct role in the biology of the individuals, including their risk of developing disease, but the virtue of the majority is that they can serve as markers for the surrounding DNA, and thus serve as leads during a search for a causative gene polymorphism, as substitutes in the evaluation of its role in health and disease, and as substitutes in the evaluation of the genetic constitution of individuals. 25

30 The association of an allele of one sequence polymorphism with particular alleles of other sequence polymorphisms in the surrounding DNA has two origins, known in the genetic field as linkage and linkage disequilibrium, respectively. Linkage arises because large parts of chromosomes are passed unchanged from parents to offspring, so that minor regions of a chromosome tend to flow unchanged from one generation to the next and also to be similar in different branches of the same family. Linkage is gradually eroded by recombination occurring in the cells of the germ-

P 687 DK01

2

line, but typically operates over multiple generations and distances of a number of million bases in the DNA.

5 Linkage disequilibrium deals with whole populations and has its origin in the (distant) forefather in whose DNA a new sequence polymorphism arose. The immediate surroundings in the DNA of the forefather will tend to stay with the new allele for many generations. Recombination and changes in the composition of the population will again erode the association, but the new allele and the alleles of any other polymorphism nearby will often be partly associated among unrelated humans even today. A
10 crude estimate suggests that alleles of sequence polymorphisms with distances less than 10000 bases in the DNA will have tended to stay together since modern man arose. Linkage disequilibrium in limited populations, for instance Europeans, often extends over longer distances. This can be the result of newer mutations, but can also be a consequence of one or more "bottlenecks" with small population sizes and
15 considerable inbreeding in the history of the current population. Two obvious possibilities for "bottlenecks" in Europeans are the exodus from Africa and the repopulation of Europe after the last ice age.

20 Linkage disequilibrium is the results of many stochastic events and as such subject to statistical variation occasionally resulting in discontinuities, lack of a monotonic relationship between association and distance and differences between people of different ethnicity. Therefore, it is often advantageous to study more than one sequence polymorphism in a given region. This also allows for further definition of the genetic surroundings of the biologically relevant polymorphism by combining the
25 associated alleles of the different markers into a so-called haplotype.

30 Humans in general carry two copies of each human chromosome in each cell. There are exceptions to this rule, not relevant to this application. We therefore speak about genotypes i.e. the combined analysis of both chromosomes at a given sequence polymorphism. The resulting genotypes of a person, analysed for instance on DNA from peripheral blood leukocytes, are inherently very stable over time. Therefore, this type of analysis can be performed any time in the life of a person and will be applicable to this person for his or her entire life. By the same token such genetic analyses are ideally suited to predict future risks of disease.

35

P 687 DK01

3

A variety of investigations suggest that many diseases in part are determined by the genetic constitution of the individual. One group of genes in particular has been associated with rare genetic predispositions to cancer. These are the genes involved in maintaining the integrity of a persons DNA, the so-called DNA repair genes. One set of such genes are the XP genes which participate in nucleotide excision repair, and, when mutated, give rise to a 1000 fold increased risk of getting skin cancer. For this reason we have previously investigated single nucleotide polymorphisms in one DNA repair gene XPD for association with risk of skin cancer in a cohort of Caucasian Americans, and found that one allele of the sequence polymorphism called XPD₆ was associated with a moderately increased risk of getting basal cell carcinoma, the most common form of skin cancer. Later other groups have studied the association between sequence polymorphisms in this and other DNA repair genes and various forms of cancer. Some have reported positive results.

Very little is known about the function of the gene RAI. It was cloned because its protein product binds to and inhibits RelA of the transcription regulator NF-kappaB.

Summary of the Invention

The present invention relates in a first aspect to a group of nucleic acid sequences found to be associated with cancer. The invention further relates to transcriptional and translational products of said sequence. An allele in the r region can be identified as correlated with an increased risk of developing cancer, the prognosis of developed cancer, and responsiveness to cancer treatment on the basis of statistical analyses of the incidence of a particular allele in individuals diagnosed with cancer.

Thus, in a first aspect the invention relates to a method for estimating the cancer risk of an individual comprising

- providing a sample from said individual,
- assessing in the genetic material including human genes in said sample a sequence polymorphism
- in a region corresponding to SEQ ID NO: 1, or a part thereof, or

P 687 DK01

4

- In a region complementary to SEQ ID NO: 1, or a part thereof, or
- In a transcription product from a sequence in a region corresponding to SEQ ID NO: 1, or a part thereof, or
- or translation product from a sequence in a region corresponding to SEQ ID NO: 1, or a part thereof,
- obtaining a sequence polymorphism response,
- estimating the cancer risk of said individual based on the sequence polymorphism response.

The estimation of the cancer risk of an individual can involve the comparison of the number and/or kind of polymorphic sequences identified with a predetermined cancer risk profile. Such a profile can be based on statistical data obtained for a relevant reference group of individuals.

The sequence of the r region is set forth as SEQ ID NO 1, originating from the cloning of human chromosome 19q published as part of the contig NT_011109 in the database of human sequences established by National Center for Biotechnology Information and located on the internet at <http://www.ncbi.nlm.nih.gov/genome/guide/human/>

The presence of an allele is determined by determining the nucleic acid sequence of all or part of the region according to standard molecular biology protocols well known in the art as described for example in Sambrook et al. (1989) and as set forth in the Examples provided herein or products of the nucleic acid sequences.

In particular, the nucleic acid molecules of the present invention represent in a first aspect nucleic acid sequences forming part of the region r corresponding to position 1522-37752 of SEQ ID NO: 1, and preferably to certain nucleic acid sequences within the gene referred to herein as RAI. As demonstrated in the Examples presented below, the RAI gene is in particular associated with human cancer diseases.

Furthermore, the invention relates to a method for estimating the cancer prognosis of an individual comprising

P 667 DK01

5

- providing a sample from said individual,

- assessing in the genetic material including human genes in said sample a sequence polymorphism

5

- in a region corresponding to SEQ ID NO: 1, or a part thereof, or
- in a region complementary to SEQ ID NO: 1, or a part thereof, or
- in a transcription product from a sequence in a region corresponding to SEQ ID NO: 1, or a part thereof, or

10

- or translation product from a sequence in a region corresponding to SEQ ID NO: 1, or a part thereof,
- obtaining a sequence polymorphism response,

15

- estimating the cancer prognosis of said individual based on the sequence polymorphism response.

20

The estimation of the cancer prognosis of an individual can involve the comparison of the number and/or kind of polymorphic sequences identified with a predetermined cancer prognosis profile. Such a profile can be based on statistical data obtained for a relevant reference group of individuals.

25

Additionally provided is a method of identifying a human subject as having an increased likelihood of responding to a treatment, comprising a) correlating the presence of an r region allele genotype with an increased likelihood of responding to treatment; and b) determining the r region allele genotype of the subject, whereby a subject having an r region allele genotype correlated with an increased likelihood of responding to treatment is identified as having an increased likelihood of responding to treatment.

30

Thus, the present invention also relates to method for estimating a treatment response of an individual suffering from cancer to a cancer treatment, comprising

- providing a sample from said individual,

P 687 DK01

6

- assessing in the genetic material including human genes in said sample a sequence polymorphism
- 5
- in a region corresponding to SEQ ID NO: 1, or a part thereof, or
 - in a region complementary to SEQ ID NO: 1, or a part thereof, or
 - in a transcription product from a sequence in a region corresponding to SEQ ID NO: 1, or a part thereof, or
 - or translation product from a sequence in a region corresponding to SEQ ID NO: 1, or a part thereof,
- 10
- obtaining a sequence polymorphism response,
- estimating the individual's response to the cancer treatment based on the sequence polymorphism response.
- 15
- The estimation of the individual's response to cancer treatment can involve the comparison of the number and/or kind of polymorphic sequences identified with a predetermined cancer treatment response profile. Such a profile can be based on statistical data obtained for a relevant reference group of individuals.
- 20
- The invention also comprises primers or probes for use in the invention, as well as kits including these. The primers and/or probes are preferably capable of hybridising to SEQ ID NO:1, or a part thereof, in particularly the r region, or a part thereof, under stringent conditions.
- 25
- Furthermore, the invention also relates to cloning vectors and expression vectors containing the nucleic acid molecules of the invention, as well as hosts which have been transformed with such nucleic acid molecules, including cells genetically engineered to contain the nucleic acid molecules of the invention, and/or cells genetically engineered to express the nucleic acid molecules of the invention. The nucleic acids are preferably isolated from the r region and preferably contain one or more sequence polymorphisms as described herein below in more detail. In addition to host cells and cell lines, hosts also include transgenic non-human animals (or progeny thereof).
- 30

P 687 DK01

7

In particular, the present invention is based on the discovery of the correlation with single nucleotide polymorphisms (SNPs) and/or tandem repeats in the r region and cancer. Thus, SNPs have been found in the r region as shown in table 1. However, the present invention is not limited to the SNPs shown in table 1, but does include any SNP in the region. Tandem repeats have been found in the r region as shown in table 2. However, the present invention is not limited to the tandem repeats shown in table 2, but does include any tandem repeat in the region.

The term human includes both a human having or suspected of having a cancer disease and an a-symptomatic human who may be tested for predisposition or susceptibility to cancer. At each position the human may be homozygous for an allele or the human may be a heterozygote.

Drawings

Fig. 1 shows a subregion of chromosome 19q

Fig. 2 shows odds ratios and p-values for individual sequence variations in relation to risk of basal cell carcinoma

Fig. 3 shows odds p-values for association of different sequence variations with risk of basal cell carcinoma among psoriatic Danes

Detailed description of the invention

The present invention relates to a characterization of a person's present and/or future risk of getting certain forms of cancer. The characterization is based on the analysis of sequence polymorphisms in a region of chromosome 19q in the person.

A number of polymorphisms in the chromosomal region 19q13.2-3 have been identified and characterised. Surprisingly, the sequence polymorphisms with strongest association to disease appeared to be located outside XPD. More specifically, the sequences were located in a sub-region between XPD and ERCC1, and seemed to have a maximum in or around the gene RAI (See Example 1). For persons getting their skin cancer relatively early (before 50 years of age), it was found that predic-

P 687 DK01

8

tions got better (Example 2) and when two sequence polymorphisms in RAI were combined, the prediction of early skin cancer got even better (Example 3). It was also possible to combine sequence polymorphisms in RAI with sequence polymorphisms outside the region and get highly positive results (Example 4).

5

The region of chromosome 19q, more precisely the region located in 19q13.2-3, with which the present invention is concerned, is depicted in Figure 1 as it is presently known together with the presently known or suspected genes. The arrows indicate the directions of transcription of the genes. The absolute chromosome positions shown are from the particular build of NCBI's map of chromosome 19, and will probably change with time.

10

The region *r* stretches from the beginning of, but not including the XPD gene, to approximately the end of ERCC1 and includes the genes RAI, LOC162978, and ASE-1. More specifically *r* is bounded by and includes the following two sequences: AGAACCCCG CCCCTCCACC TCGTCTCAA and TCCCTCCCCA GAGACTGCAC CAGCGCAGCC, and is defined by SEQ ID NO: 1.

15

In the present context the region *r* means SEQ ID NO: 1 and complementary sequence as well as transcriptional products and translational products thereof.

20

One preferred section of the region *r* stretches approximately from the end of RAI to the beginning of ASE-1 and includes the genes RAI, LOC162978, and ASE-1. More specifically, this section of *r* is bounded by and includes the following sequences: GAAGTGAGCC AAGATCACGC CACTGCACTC and GTGCCCACCT GGGCCAC-CAG AAGGTGACAC. In the present context the region *r* means SEQ ID NO: 1 bases 1522-37752 and complementary sequence as well as transcriptional products and translational products thereof.

25

Finally, in the claims the gene RAI is defined as including transcribed sequences of the gene plus a 1500 base upstream promoter region. More specifically RAI is bounded by and includes the following sequences: CATAACCACA ATGATGAGCA TGTATTGAGT and ATGTTGTCCA GGCTGGTCTT GAACTCCTGA. In the present context this section of the region *r* relates to SEQ ID NO: 1 bases 7761-22885 and

30

P 687 DK01

9

complementary sequence as well as transcriptional products and translational products thereof.

5 Modifications to the human genome map are known to occur from time to time. It is therefore possible that the defining sequences quoted above will change slightly in future maps.

10 Fragments or parts of the region *r* as used herein relates to any fragment of at least 100 nucleic acid residues in length, or multiples of 100 nucleic acid residues in length, starting from SEQ ID NO: 1 position 1, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2300, 2400, 2500, 2600, 2700, 2800, 2900, 3000, and so forth, each fragment starting position having an increment of 100 nucleic acid residues. Multiples are preferably multiples of e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49 and 50.

20 For fragments starting at position 1, the length of said fragments will thus be e.g. 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2300, 2400, 2500, 2600, 2700, 2800, 2900, 3000, and so forth, using suitable multipliers as listed herein above.

25 For fragments starting at position 100, the length of said fragments will thus be e.g. 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2300, 2400, 2500, 2600, 2700, 2800, 2900, 3000, and so forth, using suitable multipliers as listed herein above.

30 For fragments starting at position 7700, the length of said fragments will thus be e.g. 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2300, 2400, 2500, 2600, 2700, 2800, 2900, 3000, 3500, 4000, 4500, 5000, 5500, 6000, 6500, 7000, 7500, 8000, 8500, 9000, 9500, 10000, 10500, 11000, 11500, 12000, 12500, 13000, 13500, 14000, 14500, 15000, and so forth, using suitable multipliers such as e.g. the ones listed herein above.

35

P 687 DK01

10

The nucleic acid sequences according to the present invention makes it possible to estimate cancer risk in an individual by using sequence polymorphisms originating from a specific region of chromosome 19.

5 Estimation of cancer risks has a number of important applications:

10 (1) Individuals with reasons to suspect that they are at risk for getting cancer would be able to clarify their situation and, if possible, take protective action. Alternatively, anti-cancer campaigns, companies, hospitals or other institutions could offer a service to help people clarify their situation. It would for instance be possible to test persons, when they got their first basal cell carcinoma, which is often recurrent and also is a moderate predictor for other cancers. If the persons were in a high-risk group, one could then advise them about, or they could of their own accord choose, risk-reducing behaviour, such as avoidance of excessive sun-exposure, abstaining from smoking etc. About 5 percent of the Danish population will at some point in their life
15 get a basal cell carcinoma.

20 (2) Anti-cancer campaigns, companies, hospitals or other institutions would be able to define relevant target subpopulations and focus information on risk-reducing behaviour on these persons. They might perhaps also be in a position to inform the remainder of the population that they need not worry. Lung cancer affects approximately 10-15 percent of smokers and thus approximately 5 percent of the population, somewhat varying from country to country. Malignant melanoma, a sun-induced, often lethal form of skin cancer, affects approximately 700 persons a year
25 in Denmark or about 1 percent of the Danish population.

30 (3) The drugs used in cancer treatment are often carcinogenic themselves and individual responses to them vary considerably, both with respect to tolerance to the treatment and with respect to efficacy of the treatment. It is an obvious possibility that the region of chromosome 19 here dealt with, which contains DNA repair genes known to modulate carcinogen responses, also modulates response to anti-cancer agents. Hence, analysis of the region may facilitate better choices of treatment for cancer, and/or help predict the future course of disease.

P 687 DK01

11

By sequence polymorphism is understood any single nucleotide, tandem repeat, insert, deletion or block polymorphism, which varies among humans, whether it is of biological importance or not.

5

Position of sequence polymorphism in the region r

In one embodiment of the methods of the invention, preferably the method for diagnosis as described herein, one or more single nucleotide polymorphism(s) at a pre-determined position in the region r (SEQ ID NO:1) are identified and used for e.g. cancer risk profiling and/or cancer treatment response profiling. Presently preferred single nucleotide polymorphism(s) are listed in Table 1. However, the present invention relates to any SNP in the r region.

15 Table 1

	Identification in dbSNP ¹	Position in SEQ ID NO: 1
	rs#3138378 A/G	137
	rs#3138376 G/T	235
20	rs#209725 C/A ambiguous location	
	rs#2377328 C/T	7189
	rs#6966 A/T	7887 (=RAIe6)
	rs#2017154 A/C	12115
	rs#2017104 A/G	12190
25	rs#2070830 T/G	14575
	rs#1970764 A/G	15798 (=RAIi1)
	rs#2226949 T/G	32035
	rs#959457 C/T	32446
	rs#2336218 C/A	32447
30	rs#766934 A/G	32481
	rs#928911 C/T	32785
	rs#1005165 C/T	33974
	rs#1005166 C/T	34119
	rs#967591 A/G	34858 (=ASE-1e1)
35	rs#1046282 T/C	35596

P 687 DK01

12

	rs#2013521 A/T	36254
	rs#735482 A/C	36926
	rs#762562 A/G	37267
	rs#2336919 ambiguous location	
5	rs#743571 C/G	37786

¹ dbSNP is the database over SNPs established by the National Center for Biotechnology Information and located on the internet at <http://www.ncbi.nlm.nih.gov/SNP/>.

10 In another embodiment of the invention preferably the method for described herein is one in which the tandem repeat is at a position as described in Table 2:

Table 2

15 **Identification in uniSTS²**

D19S908
 STS-W67936
 D19S543
 20 D19S393
 STS-R48186
 GDB:181915
 RH47033
 GDB:190019

25 ² UniSTS is a database of unique sequence tag sites established by National Center for Biotechnology Information and located on the internet at <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=unists>

30 In another embodiment of the invention, the method for diagnosis described herein is preferably one in which the sequence polymorphism is in region r. Testing for the presence of the RAI gene allele is especially preferred because, without wishing to be bound by theoretical considerations, of its association with increased risk of cancer (as explained herein).

P 687 DK01

13

5 The sequence polymorphism of the invention comprises at least one base difference, such as at least two base differences. As described above the sequence polymorphism comprises at least one single nucleotide polymorphism, such as at least two single nucleotide polymorphisms. Also, the sequence polymorphism comprises at least one tandem repeat polymorphism, such as at least two tandem repeat polymorphisms.

10 Also, the sequence polymorphism may be a combination of single nucleotide polymorphism and tandem repeats.

The status of the individual may be determined by reference to allelic variation at one, two, three, four or more of the above loci.

15 Cell sample

The cell sample used in the present invention may be any suitable cell sample capable of providing the genetic material for use in the method. In a preferred embodiment, the cell sample is a blood sample, a tissue sample, a sample of secretion, semen, ovum, a washing of a body surface (e.g. a buccal swap), a clipping of a body surface (hairs, or nails), such as wherein the cell is selected from white blood cells and tumour tissue.

20 It will be appreciated that the test sample may equally be a nucleic acid sequence corresponding to the sequence in the test sample, that is to say that all or a part of the region in the sample nucleic acid may firstly be amplified using any convenient technique e.g. PCR, before use in the analysis of variation in the region.

Detection methods

30 Detection may be conducted on the sequence of SEQ ID NO: 1 or a complementary sequence as well as on translational (mRNA) and transcriptional products (polypeptides, proteins) therefrom.

35 It will be apparent to the person skilled in the art that there are a large number of analytical procedures which may be used to detect the presence or absence of vari-

P 687 DK01

14

ant nucleotides at one or more of positions mentioned herein in the r region. Mutations or polymorphisms within or flanking the r region can be detected by utilizing a number of techniques. Nucleic acid from any nucleated cell can be used as the starting point for such assay techniques, and may be isolated according to standard nucleic acid preparation procedures that are well known to those of skill in the art. In general, the detection of allelic variation requires a mutation discrimination technique, optionally an amplification reaction and a signal generation system. Table 3 lists a number of mutation detection techniques, some based on the PCR. These may be used in combination with a number of signal generation systems, a selection of which is listed in Table 4. Further amplification techniques are listed in Table 5. Many current methods for the detection of allelic variation are reviewed by Nollau et al., Clin. Chem. 43, 1114-1120, 1997; and in standard textbooks, for example "Laboratory Protocols for Mutation Detection", Ed. by U. Landegren, Oxford University Press, 1996 and "PCR", 2.sup.nd Edition by Newton & Graham, BIOS Scientific Publishers Limited, 1997.

Table 3

Abbreviations:		
20	ALEX .TM.	Amplification refractory mutation system linear extension
	APEX	Arrayed primer extension
	ARMS .TM.	Amplification refractory mutation system
	b-DNA	Branched DNA
	CMC	Chemical mismatch cleavage
25	bp	base pair
	COPS	Competitive oligonucleotide priming system
	DGGE	Denaturing gradient gel electrophoresis
	FRET	Fluorescence resonance energy transfer
	LCR	Ligase chain reaction
30	MASDA	Multiple allele specific diagnostic assay
	NASBA	Nucleic acid sequence based amplification
	OLA	Oligonucleotide ligation assay
	PCR	Polymerase chain reaction
	PTT	Protein truncation test
35	RFLP	Restriction fragment length polymorphism

P 887 DK01

15

	SDA	Strand displacement amplification
	SNP	Single nucleotide polymorphism
	SSCP	Single-strand conformation polymorphism analysis
	SSR	Self sustained replication
5	TGGE	Temperature gradient gel electrophoresis

Table 4 illustrates various mutation detection techniques capable of being used for SNP detection.

10

Table 4

General techniques: DNA sequencing, Sequencing by hybridisation, SNAPshot.

15 Scanning techniques: PJT*, SSCP, DOGE, TGGE, Cleavase, Heteroduplex analysis, CMC, Enzymatic mismatch cleavage

Hybridisation Based techniques

20 Solid phase hybridisation: Dot blots, MASDA, Reverse dot blots, Oligonucleotide arrays (DNA Chips)

25 Solution phase hybridisation: Taqman.TM.--U.S. Pat. No. 5,210,015 & 5,487,972 (Hoffmann-La Roche), Molecular Beacons--Tyagi et al (1996), Nature Biotechnology, 14, 303; WO 95/13399 (Public Health Inst., New York), Lightcycler, optionally in combination with FRET.

30 Extension Based: ARMS.TM., ALEX.TM.--European Patent No. EP 332435 B1 (Zeneca Limited), COPS--Gibbs et al (1989), Nucleic Acids Research, 17, 2347.

Incorporation Based: Mini-sequencing, APEX

Restriction Enzyme Based: RFLP, Restriction site generating PCR

35 Ligation Based: OLA

P 687 DK01

16

Other: Invader assay

Various Signal Generation or Detection Systems is listed below:

5

Fluorescence: FRET, Fluorescence quenching, Fluorescence polarisation—United Kingdom Patent No. 2228998 (Zeneca Limited)

10

Other: Chemiluminescence, Electrochemiluminescence, Raman, Radioactivity, Colorimetric, Hybridisation protection assay, Mass spectrometry

Table 5 illustrates examples of further amplification techniques.

15

Table 5

SSR, NASBA, LCR, SDA, b-DNA

20

Preferred mutation detection techniques include ARMS.TM., ALEX.TM., COPS, Taqman, Molecular Beacons, RFLP, and restriction site based PCR and FRET techniques.

25

Particularly preferred methods include FRET; taqman, ARMS.TM. and RFLP based methods.

30

In a preferred embodiment, mutations or polymorphisms can be detected by using a microassay of nucleic acid sequences immobilized to a substrate or "gene chip" (see, e.g. Cronin, et al., 1996, Human Mutation 7:244-255).

35

Further, improved methods for analyzing DNA polymorphisms, which can be utilized for the identification of region r specific mutations, have been described that capitalize on the presence of variable numbers of short, tandemly repeated DNA sequences between the restriction enzyme sites. For example, Weber (U.S. Pat. No. 5,075,217) describes a DNA marker based on length polymorphisms in blocks of (dC-dA)_n-(dG-dT)_n short tandem repeats. The average separation of (dC-dA)_n-(dG-

P 687 DK01

17

dT)n blocks is estimated to be 30,000-60,000 bp. Markers that are so closely spaced exhibit a high frequency co-inheritance, and are extremely useful in the identification of genetic mutations, such as, for example, mutations within the RAI gene, and the diagnosis of diseases and disorders related to RAI mutations.

5

Also, Caskey et al. (U.S. Pat. No. 5,364,759) describe a DNA profiling assay for detecting short tri and tetra nucleotide repeat sequences. The process includes extracting the DNA of interest, such as the RAI gene, amplifying the extracted DNA, and labelling the repeat sequences to form a genotypic map of the individual's DNA.

10

The level of RAI gene expression can also be assayed. For example, RNA from a cell type or tissue known, or suspected, to express the RAI gene, such as brain, may be isolated and tested utilizing hybridization or PCR techniques such as are described, above. The isolated cells can be derived from cell culture or from a patient. The analysis of cells taken from culture may be a necessary step in the assessment of cells to be used as part of a cell-based gene therapy technique or, alternatively, to test the effect of compounds on the expression of the RAI gene. Such analyses may reveal both quantitative and qualitative aspects of the expression pattern of the RAI gene, including activation or inactivation of RAI gene expression.

15

20

In one embodiment of such a detection scheme, a cDNA molecule is synthesized from an RNA molecule of interest (e.g., by reverse transcription of the RNA molecule into cDNA). A sequence within the cDNA is then used as the template for a nucleic acid amplification reaction, such as a PCR amplification reaction, or the like.

25

The nucleic acid reagents used as synthesis initiation reagents (e.g., primers) in the reverse transcription and nucleic acid amplification steps of this method are chosen from among the RAI gene nucleic acid reagents described above. The preferred lengths of such nucleic acid reagents are at least 9-30 nucleotides. For detection of the amplified product, the nucleic acid amplification may be performed using radioactively or non-radioactively labeled nucleotides. Alternatively, enough amplified product may be made such that the product may be visualized by standard ethidium bromide staining or by utilizing any other suitable nucleic acid staining method.

30

Additionally, it is possible to perform such RAI gene expression assays "in situ", i.e., directly upon tissue sections (fixed and/or frozen) of patient tissue obtained from

35

P 687 DK01

18

biopsies or resections, such that no nucleic acid purification is necessary. Nucleic acid reagents such as those described above may be used as probes and/or primers for such in situ procedures (see, for example, Nuovo, G. J., 1992, "PCR In Situ Hybridization: Protocols And Applications", Raven Press, NY).

5

Alternatively, if a sufficient quantity of the appropriate cells can be obtained, standard Northern analysis can be performed to determine the level of mRNA expression of the RAI gene.

10 Activity of the gene

Another method for detecting sequence polymorphism is by analysing the activity of gene products resulting from the sequences. Accordingly, in one embodiment the detection uses the activity of the RAI gene product as compared to a reference in the method. In particular if the activity of the genes are decreased or increased by at least or about 50 %, such as at least or about 40%, for example at least or about 30%, such as at least or about 20%, for example at least or about 10%, such as at least or about 10%, for example at least or about 5%, such as at least or about 2%, it indicates a sequence polymorphism in the gene.

20

Mutations outside the region

The present invention may combine the result of sequence polymorphism within the region r with sequence polymorphism outside the region in order to increase the probability of the correlation.

25

Primers

The primers nucleotide sequences of the invention further include: (a) any nucleotide sequence that hybridizes to a nucleic acid molecule of the region r or its complementary sequence or RNA products under stringent conditions, e.g., hybridization to filter-bound DNA in 6x sodium chloride/sodium citrate (SSC) at about 45°C followed by one or more washes in 0.2x SSC/0.1% SDS at about 50-65°C, or (b) under highly stringent conditions, e.g., hybridization to filter-bound nucleic acid in 6x SSC at about 45°C followed by one or more washes in 0.1x SSC/0.2% SDS at about

30

35

P 687 DK01

19

68°C, or under other hybridization conditions which are apparent to those of skill in the art (see, for example, Ausubel F.M. et al., eds., 1989, *Current Protocols in Molecular Biology*, Vol. 1, Green Publishing Associates, Inc., and John Wiley & sons, Inc., New York, at pp. 6.3.1-6.3.6 and 2.10.3). Preferably the nucleic acid molecule that hybridizes to the nucleotide sequence of (a) and (b), above, is one that comprises the complement of a nucleic acid molecule of the region r or a complementary sequence or RNA product thereof. In a preferred embodiment, nucleic acid molecules comprising the nucleotide sequences of (a) and (b), comprises nucleic acid molecule of RAI or a complementary sequence or RNA product thereof.

10

Among the nucleic acid molecules of the invention are deoxyoligonucleotides ("oligos") which hybridize under highly stringent or stringent conditions to the nucleic acid molecules described above. In general, for probes between 14 and 70 nucleotides in length the melting temperature (T_m) is calculated using the formula:

15

$$T_m(^{\circ}\text{C}) = 81.5 + 16.6(\log [\text{monovalent cations (molar)}]) + 0.41(\% \text{ G+C}) - (500/N)$$

where N is the length of the probe. If the hybridization is carried out in a solution containing formamide, the melting temperature is calculated using the equation $T_m(^{\circ}\text{C}) = 81.5 + 16.6(\log [\text{monovalent cations (molar)}]) + 0.41(\% \text{ G+C}) - (0.61\% \text{ formamide}) - (500/N)$ where N is the length of the probe. In general, hybridization is carried out at about 20-25 degrees below T_m (for DNA-DNA hybrids) or 10-15 degrees below T_m (for RNA-DNA hybrids).

25

Exemplary highly stringent conditions may refer, e.g., to washing in 6x SSC/0.05% sodium pyrophosphate at 37°C (for about 14-base oligos), 48°C (for about 17-base oligos), 55°C (for about 20-base oligos), and 60°C (for about 23-base oligos).

30

Accordingly, the invention further provides nucleotide primers or probes which detect the r region polymorphisms of the invention. The assessment may be conducted by means of at least one nucleic acid primer or probe, such as a primer or probe of DNA, RNA or a nucleic acid analogue such as peptide nucleic acid (PNA) or locked nucleic acid (LNA). The nucleotide primer or probe is preferably capable of hybridizing to a subsequence of the region corresponding to SEQ ID NO: 1, or a part thereof, or a region complementary to SEQ ID NO:1.

35

P 687 DK01

20

According to one aspect of the present invention there is provided an allele-specific oligonucleotide probe capable of detecting a r region polymorphism at one or more of positions in the r region as defined by the positions in SEQ ID NO: 1.

5

The allele-specific oligonucleotide probe is preferably 5-50 nucleotides, more preferably about 5-35 nucleotides, more preferably about 5-30 nucleotides, more preferably at least 9 nucleotides.

10

The design of such probes will be apparent to the molecular biologist of ordinary skill. Such probes are of any convenient length such as up to 50 bases, up to 40 bases, more conveniently up to 30 bases in length, such as for example 8-25 or 8-15 bases in length. In general such probes will comprise base sequences entirely complementary to the corresponding wild type or variant locus in the region. However, if required one or more mismatches may be introduced, provided that the discriminatory power of the oligonucleotide probe is not unduly affected. The probes of the invention may carry one or more labels to facilitate detection.

15

20

In one embodiment, the primers and/or probes are capable of hybridizing to a subsequence selected from the group of subsequences below, wherein the polymorphism is denoted as for example T/C:

25

30

35

1. GCTCTGAAAC TTACTAGCCC(A/G)GTATTTATGG AGAGGCATTT
2. GTGGTCAAAT TCTCATTCAT CGTGG (T/C) CCAGGCAAGC
ACACTTCCTC
3. ACCCTGAGGT GAGCACCTGT TCCTT(C/T) TCCTTGCCCT TAGCCCA-
GAG GTAGA
4. GGGCAGGGGT TTGTGCCTCC AATGA (G/A) CACAAGCTCC
CCCTGCCCCC CAACT
5. CCTGGCGGTG GCCGTCACCA GCTTT (T/C) GGGGGTGTTT
GGGAAGCTGG
6. CTCCAGCCCC ACTGTTCCCT (A/G) GGCCCTATTG GTCCCCCTGG
7. ACAAGGAGGA GGCAGAAGTG AGGTT (G/C) AAACCCACTG CCCAATC-
TTA
8. CCAACACGGT GAAACCCCGT CTGTA(T/C)TAAAAATACA AAAATTAGCC

P 687 DK01

21

9. AATCCAGGAC CCCATAATCT TCCGT (C/T) ATCTAAAACA ATA-
ATGGTGA
10. CCCAAGGGGG CGAGGGGAGG GTGAA (A/G)GGGTGGGACG
GGGGCAGCCG
- 5 11. GAAGTGAGAA GGGGGCTGGG GGTCG (G/-) CGCTCGCTAG
CGGGCGCGGG
12. CGCACGCGCA GTATCCCGAT TGGCT (C/G)TGCCCTAGCG GATT-
GACGGG
13. AACTCCTGGG TTCGATCAAT ACTCA (GACA/-) ATCTTGGCAG
GCGCAGGAGG
- 10 14. GCTGGGATTA CAGGCTTGAG CCACC (A/G) CGCCCGGCCT
GCAAAGCCAT
15. TTTTGTATCT TTAGTAGAGA CAGG (T/G) TTTCTCCATG TTGGTCAGGC
16. GCCTCAGCCT CCCGAGTAGC TGAGACT (C/A) CAGGTGCCCCG CCAC-
CACGCC
- 15 17. TGAAATTGTA GGTTGAGAGG CCAGGCG (C/T) GGTGCTCACG
CCTGTAATTT
18. GTTTATAAAC ATTAAACCAG (T/A) GCTGTGTGAA GGCACCTTAAT
19. CCGTCTCTAT TAAAAATATA AAA (A/C) AATTTAGCCG GGTGTAGCGG
- 20 20. GGGAGGCTCG AGGCGGGC (A/G) GATTGCATGA GCTCAGGATT
21. TCCCAAGTTT CAGGGCCCAA (T/G) ATTCTCAAAT CACAGGATTC
22. TGCAGTGAGC TGAGATCGC (A/G) CCACTGCACT CCAGCCTGGG
23. TCTTAGGACG CATGGGGGT (T/G) GAGAGAACGG GGAGATAGAC
24. CTGGGTTCTA GAACTACC (C/T) ATGCAAACCC AGCTGTTTCC
- 25 25. ATTCTGCCCT GGGTTCTAGA ACTACCT (C/A) TGCAAACCCA
GCTGTTTCCC
26. GCTGTTTCCC ACCCCATAAG GCA (A/G) TAGGGGAGCC
CACCTCCGCC
27. GACCTAGAAG ATCGGTCGAG A (C/T) AGCAGCTTGA GGCTGGCAGG
- 30 28. CTGGCCAGGA ATGCAGTCGG GTCAC (C/T) CTGTCTAGCC
ACCGTCTCGC
29. GGGAGGAGTC GCCGATCAGG (C/T) CCCTTCCTGA AAGTCATCGA
30. GCAGCCCCGGG CTACAGGGTT (A/G) CCTGAGGTGT GGGTCCCAGG
31. TAGAAATACT AACAAAGGGC (T/C) GTGGGTTTCT CCCCCTGCTT
- 35 32. ACAGGAGAGG GAAGGTTTTTTG (A/T) TTTTTTTTTT GTTTTTTTTT

P 687 DK01

22

33. GAAGAGGAAG AAGCCCAAAG GGA (A/C) AGAAACCTTC GAGCCA-
GAAG
34. GCGCCTCAAC AGCCAGAAGG AGCG (A/G) AGCCTCAGGC CCAGG-
CAGCT
- 5 35. TTGAGACTCT CTGTTTGAT (A/G) CTTCACTCAG AAGGTGCTTC
36. AGGCCAGGCT CCTGCTGGCT G (C/G) GCTGGTGCAG TCTCTGGGGA
37. CCCCTATACC CTCAAGCAT (C/T) TATCCATTGA GTTACAAACA
38. ACCATCCCCC GCCTTCCGTT (A/C) GTCCGGCCCC CGAGGCTAGC

10

In another embodiment, the primers and/or probes are capable of hybridizing to a subsequence selected from the group of subsequences below:

- 15 1. TGAAATTGTA GGTGAGAGG CCAGGCG (C/T) GGTGCTCACG
CCTGTAATT
2. GTTTATAAAC ATTAACCAG (T/A) GCTGTGTGAA GGCACCTAAT
3. CCGTCTCTAT TAAAAATATA AAA (A/C) AATTAGCCG GGTGTAGCGG
4. GGGAGGCTCG AGGCGGGC (A/G) GATTGCATGA GCTCAGGATT
- 20 5. TCCCAAGTTT CAGGGCCCAA (T/G) ATTCTCAAAT CACAGGATT
6. TGCAGTGAGC TGAGATCGC (A/G) CCACTGCACT CCAGCCTGGG
7. TCTTAGGACG CATGGGGGT (T/G) GAGAGAACGG GGAGATAGAC
8. CTGGGTTCTA GAACTACC (C/T) ATGCAAACCC AGCTGTTTCC
9. ATTCTGCCCT GGGTTCTAGA ACTACCT (C/A) TGCAAACCCA
GCTGTTTCCC
- 25 10. GCTGTTTCCC ACCCCATAAG GCA (A/G) TAGGGGAGCC
CACCTCCGCC
11. GACCTAGAAG ATCGGTCGAG A (C/T) AGCAGCTTGA GGCTGGCAGG
12. CTGGCCAGGA ATGCAGTCGG GTCAC (C/T) CTGTCTAGCC
ACCGTCTCGC
- 30 13. GGGAGGAGTC GCCGATCAGG (C/T) CCCTTCCTGA AAGTCATCGA
14. GCAGCCCGGG CTACAGGGTT (A/G) CCTGAGGTGT GGGTCCCAGG
15. TAGAAATACT AACAAAGGGC (T/C) GTGGGTTTCT CCCCCTGCTT
16. ACAGGAGAGG GAAGGTTTTTTG (A/T) TTTTTTTTTT GTTTTTTTTT
- 35 17. GAAGAGGAAG AAGCCCAAAG GGA (A/C) AGAAACCTTC GAGCCA-
GAAG

P 687 DK01

23

18. GCGCCTCAAC AGCCAGAAGG AGCG (A/G) AGCCTCAGGC CCAGG-CAGCT

5 In yet another embodiment, the primers and/or probes are capable of hybridizing to a subsequence selected from the group of subsequences below

1. GTTTATAAAC ATTAAACCAG (T/A) GCTGTGTGAA GGCACTTAAT
2. CCGTCTCTAT TAAAAATATA AAA (A/C) AATTTAGCCG GGTGTAGCGG
3. GGGAGGCTCG AGGCGGGC (A/G) GATTGCATGA GCTCAGGATT
- 10 4. TCCCAAGTTT CAGGGCCCAA (T/G) ATTCTCAAAT CACAGGATTC
5. TGCAGTGAGC TGAGATCGC (A/G) CCACTGCACT CCAGCCTGGG

15 It is preferred in one embodiment that at least one sequence polymorphism is assessed in a region corresponding to SEQ ID NO: 1 position 1521-37752 (r), including at least one sequence polymorphism assessed in a region corresponding to SEQ ID NO: 1 position 7760-22885.

20 In another embodiment, the methods of the invention relates to at least one sequence polymorphism is assessed in a region corresponding to SEQ ID NO: 1 position 34391-37683, ending with the coding region of ASE-1 (cagcctgttag), where tag is the stop codon.

25 In a preferred embodiment the primers or probes are selected from one or more of the following:

- 26 TGGCTAACACGGTGAAACC (SEQ ID NO:7)
- 27 GGAATCCAAAGATTCTATGATGG (SEQ ID NO:8)
- 28 GGGAGGCGGAGCTTGCACTGA (SEQ ID NO:9)
- 29 CTGAGATCGCACCACTGCAC (SEQ ID NO:10)
- 30 GGTTTTCTGCTCTGCACACG (SEQ ID NO:11)
- 31 CCTTTCTCCTTCCACCAACG (SEQ ID NO:12)
- 32 CGGGCTACAGGGTTACCTGAG (SEQ ID NO:13)
- 33 TCTGCAACCTGGTGCGAGCAGC (SEQ ID NO:14)
- 34 CCTACCACCATCATCACATCC (SEQ ID NO:15)

P 687 DK01

24

GCCTTGCCAAAAATCATAACC (SEQ ID NO:16)

CCTCTCCCCAATTAAGTGCCTTCACACAGC (SEQ ID NO:17)

AGCCAGGGAGGTTGAGGCT (SEQ ID NO:18)

AGACAGCCCTGAATCAGCAC (SEQ ID NO:19)

5 GCAATGAGCCGAGATAGAA (SEQ ID NO:20)

TGGCTAGCCCATTACTCTA (SEQ ID NO:21)

10 According to another aspect of the present invention there is provided a diagnostic nucleic acid primer capable of detecting a *r* region polymorphism at one or more of positions in the *r* region as defined by the In SEQ ID NO: 1.

15 The primer or probe may be a diagnostic nucleic acid primer defined as an allele specific primer, used, generally together with a constant primer, in an amplification reaction such as a PCR reaction, which provides the discrimination between alleles through selective amplification of one allele at a particular sequence position. The diagnostic primer is preferably 5-50 nucleotides, more preferably about 5-35 nucleotides, more preferably about 5-30 nucleotides, more preferably at least 9 nucleotides.

20 In accordance with the present invention diagnostic primers are provided, comprising the sequences set out below as well as derivatives thereof wherein about 6-8 of the nucleotides at the 3' terminus are identical to the sequences given below and wherein up to 10, such as up to 8, 6, 4, 2, or 1 of the remaining nucleotides may be varied without significantly affecting the properties of the diagnostic primer. Conveniently, the sequence of the diagnostic primer is as written below.

25

30 Furthermore, as described above at least two sets of primer(s) and/or probe(s) may be combined in the method thereby increasing the correlation probability. This second or other set of primer(s) and/or probe(s) may be a nucleotide or nucleotide analogues hybridising to a region within the region *r* or to a sequence different from the region *r*. Said sequence different from the region *r* is preferably a region in chromosome 19, preferably in chromosome 19q. In particular such second or other primer or probe may be selected from one or more of the sequences below, or the complementary strands:

P 887 DK01

25

GCCCCGTCCCAGGTA (SEQ ID NO:21)
AGCCCCAAGACCCCTTCACT (SEQ ID NO:22)
GTCCCATAGATAGGAGTGAAAG (SEQ ID NO:23)
5 CCCTAGGACACAGGAGCACA (SEQ ID NO:24)
TTGTGCTTTCTCTGTGTCCA (SEQ ID NO:25)
TATCAGAAAAGGCTGGAGGA (SEQ ID NO:26)
GAGTGGCTGGGGAGTAGGA (SEQ ID NO:27)
GCCAAGCAGAAGAGACAAA (SEQ ID NO:28)
10 CCTCAGATGTCCTCTGCTCA (SEQ ID NO:29)
GCCACAGCCCCAGCAAGTAG (SEQ ID NO:30)
AGGACCACAGGACACGCAGA (SEQ ID NO:31)
CATAGAACAGTCCAGAACAC (SEQ ID NO:32)
TTAGCTTGGCACGGCTGTCCAAGGA (SEQ ID NO:33)
15 ACAGAATTCGCCCCGGCCTGGTACAC (SEQ ID NO:34)
TTGAACTGGAAGTCTGAGAAGG (SEQ ID NO:35)
TGGTGGATGGTGTGAAGCA (SEQ ID NO:36)
CCTTTCTCCAACTTCTTCTCCATTTCACC (SEQ ID NO:37)
GGGGATCATGTCGTCAATGGACT (SEQ ID NO:38)
20 ATGCCCTGTAGGTTCAATGG (SEQ ID NO:39)
TGGAGGTCTTTAGGGGCTTG (SEQ ID NO:40)
GGCTGGTCCCCGTCTTCTCCTTCC (SEQ ID NO:41)
TCTCTGTTGCCACTTCAGCCTC (SEQ ID NO:42)
GTCCTGCCCTCAGCAAAGAGAA (SEQ ID NO:43)
25 TTCTCCTGCGATTAAAGGCTGT (SEQ ID NO:44)
ATCCTGTCCCTACTGGCCATTC (SEQ ID NO:45)
TGTGGACGTGACAGTGAGAAAT (SEQ ID NO:46)
TGGAGTGCTATGGCACGATCTCT (SEQ ID NO:47)
CCATGGGCATCAAATTCCTGGGA (SEQ ID NO:48)
30 CACACCTGGCTCATTTTTGTAT (SEQ ID NO:49)
TCATCCAGGTTGTAGATGCCA (SEQ ID NO:50)
AGGCTCAACAAGGAAAAATGC (SEQ ID NO:51)
GCTAGACAGTCAAGGAGGGACG (SEQ ID NO:52)
AAAGGGTGGGTGTGGGAGACATTGG (SEQ ID NO:53)
35 AAACCAACCTAGGCACCCCAA (SEQ ID NO:54)

P 687 DK01

28

CAGTGTCCAAAGAGCACC (SEQ ID NO:55)
CTACCCCTTTAGCGACC (SEQ ID NO:56)
TCCTGCCCCCAGAGCGTCACC (SEQ ID NO:57)
GTACGGTCCACATAATTTTGGAGGA (SEQ ID NO:58)
5 CGACGAACCTTCTCTGAAGCGAA (SEQ ID NO:59)
AGCGACACGGGCATCTGG (SEQ ID NO:60)
ATGAGCGTCCACCTCCTGAACC (SEQ ID NO:61)
AGGCAGCAGCATCGTCATCCCC (SEQ ID NO:62)
TGCATAGCTAGGTCCTGC (SEQ ID NO:63)
10 AACTGACRAAACTAGCTCTATGGGGTGGTGCCGCA (SEQ ID NO:64)
CTGGCTCTGAACTTACTAGCCC (SEQ ID NO:65)
GCTGGACTGTCACCGCATG (SEQ ID NO:66)
GGAGCAGGGTTGGCGTG (SEQ ID NO:67)
TGCCCTCCCAGAGGTAAGGCCT (SEQ ID NO:68)
15 CCCTCCCGGAGGTAAGGCCTC (SEQ ID NO:69)
GATCAAAGAGACAGACGAGC (SEQ ID NO:70)
GAAGCCCAGGAAATGC (SEQ ID NO:71)
GGACGCCACCTGGCCAACC (SEQ ID NO:72)
CGTGCTGCCCAACGAAGTG (SEQ ID NO:73)
20

The primers and probes may be manufactured using any convenient method of synthesis. Examples of such methods may be found in standard textbooks, for example "Protocols for Oligonucleotides and Analogues; Synthesis and Properties,"
25 Methods in Molecular Biology Series; Volume 20; Ed. Sudhir Agrawal, Humana ISBN: 0-89603-247-7; 1993; 1st Edition. If required the primer(s) and probe(s) may be labelled to facilitate detection.

Kit

30

According to another aspect of the present invention, there is provided a diagnostic kit comprising at least one diagnostic primer of the invention and/or at least one allele-specific oligonucleotide primer of the invention.

P 687 DK01

27

The diagnostic kits may comprise appropriate packaging and instructions for use in the methods of the invention. Such kits may further comprise appropriate buffer(s) and polymerase(s) such as thermostable polymerases, for example taq polymerase.

- 5 Preferred kits can comprise means for amplifying the relevant sequence such as primers, polymerase, deoxynucleotides, buffer, metal ions; and/or means for discriminating the polymorphism, such as one or a set of probes hybridising to the polymorphic site, a sequence reaction covering the polymorphic site, an enzyme or an antibody; and/or a secondary amplification system, such as enzyme-conjugated
10 antibodies, or fluorescent antibodies. The kit-of-parts preferably also comprises a detection system, such as a fluorometer, a film, an enzyme reagent or another highly sensitive detection device.

- 15 The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits. The invention therefore also encompasses kits for detecting the presence of a polypeptide or nucleic acid of the invention in a biological sample (i.e., a test sample). Such kits can be used, e.g., to determine if a subject is suffering from or is at increased risk of developing a disorder associated with a disorder-causing allele, or aberrant expression or activity of a polypeptide of the invention.
20 For example, the kit can comprise a labeled compound or agent capable of detecting the polypeptide or mRNA or DNA or RAI gene sequences, e.g., encoding the polypeptide in a biological sample. The kit can further comprise a means for determining the amount of the polypeptide or mRNA in the sample (e.g., an antibody which binds the polypeptide or an oligonucleotide probe which binds to DNA or
25 mRNA encoding the polypeptide). Kits can also include instructions for observing that the tested subject is suffering from or is at risk of developing a disorder associated with aberrant expression of the polypeptide if the amount of the polypeptide or mRNA encoding the polypeptide is above or below a normal level, or if the DNA correlates with presence of an RAI allele that causes a disorder.

- 30 For antibody-based kits, the kit can comprise, for example: (1) a first antibody (e.g., attached to a solid support) which binds to a polypeptide of the invention; and, optionally, (2) a second, different antibody which binds to either the polypeptide or to the first antibody and is conjugated to a detectable agent.
35

P 887 DK01

28

Identification of an allele as having implication for risk of cancer

5 An allele in the r region can be identified as correlated with an increased risk of developing cancer on the basis of statistical analyses of the incidence of a particular allele in two groups of individuals with and without cancer, respectively, according to the χ^2 test, which is well known in the art. Furthermore, an allele in the region can be identified as an allele correlated with prognosis of cancer on the basis of statistical analyses of the incidence of a particular allele in individuals demonstrating different prognostic characteristics.

10

Identification of humans having increased likelihood of responding to treatment

15 It is further contemplated that the present invention provides a method for identifying a human subject as having an increased likelihood of responding positively to a cancer treatment, comprising determining the presence in the subject of a r region allele genotype correlated with an increased likelihood of positive response to treatment, whereby the presence of the genotype identifies the subject as having an increased likelihood of responding to cancer treatment.

20

The treatment mentioned herein may be any cancer treatment, such as conventional cancer treatment, for example X-ray, chemotherapeutics, surgical excision or combinations thereof.

25

Protein Products of the Gene(s)

Gene products of the region r or peptide fragments thereof, can be prepared for a variety of uses. For example, such gene products, or peptide fragments thereof, can be used for the generation of antibodies, in diagnostic assays.

30

The gene products of the invention include, but are not limited to, human RAI gene products, and ASE-1 gene products. In the following the invention is described in relation to RAI gene product.

P 687 DK01

29

Gene product, sometimes referred to herein as an "protein" or "polypeptide", includes those gene products encoded by the RAI gene sequences shown as position 7821-21350 in SEQ ID NO: 1. Among gene product variants are gene products comprising amino acid residues encoded by the polymorphisms. Such gene product variants also include a variant of the RAI gene product.

In addition, RAI gene products may include proteins that represent functionally equivalent gene products. In preferred embodiments, such functionally equivalent RAI gene products are naturally occurring gene products. Functionally equivalent RAI gene products also include gene products that retain at least one of the biological activities of the RAI gene products described above, and/or which are recognized by and bind to antibodies (polyclonal or monoclonal) directed against RAI gene products.

Antibodies to Gene Products

Described herein are methods for the production of antibodies capable of specifically recognizing one or more gene product epitopes or epitopes of conserved variants or peptide fragments of the gene products. Furthermore, antibodies that specifically recognize mutant forms are encompassed by the invention. The terms "specifically bind" and "specifically recognize" refer to antibodies that bind to RAI gene product epitopes at a higher affinity than they bind to non-RAI (e.g., random) epitopes.

Such antibodies may include, but are not limited to, polyclonal antibodies, monoclonal antibodies (mAbs), humanized or chimeric antibodies, single chain antibodies, Fab fragments, F(ab')₂ fragments, fragments produced by a Fab expression library, anti-Idiotypic (anti-Id) antibodies, and epitope-binding fragments of any of the above, including the polyclonal and monoclonal antibodies described below. Such antibodies may be used, for example, in the detection of a gene product in an biological sample and may, therefore, be utilized as part of a diagnostic or prognostic technique whereby patients may be tested for abnormal levels of gene products, and/or for the presence of abnormal forms of such gene products. Such antibodies may also be utilized in conjunction with, for example, compound screening schemes, as

P 687 DK01

30

described, below, for the evaluation of the effect of test compounds on gene product levels and/or activity.

5 For the production of antibodies against a gene product, various host animals may be immunized by injection with a RAI gene product, or a portion thereof. Such host animals may include, but are not limited to rabbits, mice, and rats, to name but a few. Various adjuvants may be used to increase the immunological response, depending on the host species, including but not limited to Freund's (complete and incomplete), mineral gels such as aluminum hydroxide, surface active substances
10 such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanin, dinitrophenol, and potentially useful human adjuvants such as BCG (bacille Calmette-Guerin) and *Corynebacterium parvum*.

15 Polyclonal antibodies are heterogeneous populations of antibody molecules derived from the sera of animals immunized with an antigen, such as a gene product, or an antigenic functional derivative thereof. For the production of polyclonal antibodies, host animals such as those described above, may be immunized by injection with gene product supplemented with adjuvants as also described above.

20 Monoclonal antibodies, which are homogeneous populations of antibodies to a particular antigen, may be obtained by any technique that provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique of Kohler and Milstein, (1975, *Nature* 256:495-497; and U.S. Pat. No. 4,376,110), the human B-cell hybridoma technique (Kosbor
25 et al., 1983, *Immunology Today* 4:72; Cole et al., 1983, *Proc. Natl. Acad. Sci. U.S.A.* 80:2026-2030), and the EBV-hybridoma technique (Cole et al., 1985, *Monoclonal Antibodies And Cancer Therapy*, Alan R. Liss, Inc., pp. 77-96). Such antibodies may be of any immunoglobulin class including IgG, IgM, IgE, IgA, IgD and any subclass thereof. The hybridoma producing the mAb of this invention may be cultivated in
30 vitro or in vivo. Production of high titers of mAbs in vivo makes this the presently preferred method of production.

In addition, techniques developed for the production of "chimeric antibodies" (Morrison, et al., 1984, *Proc. Natl. Acad. Sci.*, 81:6851-6855; Neuberger, et al., 1984, *Nature* 312:604-608; Takeda, et al., 1985, *Nature*, 314:452-454) by splicing the genes
35

P 687 DK01

31

5 from a mouse antibody molecule of appropriate antigen specificity together with genes from a human antibody molecule of appropriate biological activity can be used. A chimeric antibody is a molecule in which different portions are derived from different animal species, such as those having a variable region derived from a murine mAb and a human immunoglobulin constant region. (See, e.g., Cabilly et al., U.S. Pat. No. 4,816,567; and Boss et al., U.S. Pat. No. 4,816,397, which are incorporated herein by reference in their entirety.)

10 In addition, techniques have been developed for the production of humanized antibodies. (See, e.g., Queen, U.S. Pat. No. 5,585,089, which is incorporated herein by reference in its entirety.) An immunoglobulin light or heavy chain variable region consists of a "framework" region interrupted by three hypervariable regions, referred to as complementarily determining regions (CDRs). The extent of the framework region and CDRs have been precisely defined (see, "Sequences of Proteins of Immunological Interest", Kabat, E. et al., U.S. Department of Health and Human Services (1983)). Briefly, humanized antibodies are antibody molecules from non-human species having one or more CDRs from the non-human species and a framework region from a human immunoglobulin molecule.

20 Alternatively, techniques described for the production of single chain antibodies (U.S. Pat. No. 4,946,778; Bird, 1988, Science 242:423-426; Huston, et al., 1988, Proc. Natl. Acad. Sci. U.S.A. 85:5879-5883; and Ward, et al., 1989, Nature 334:544-546) can be adapted to produce single chain antibodies against gene products. Single chain antibodies are formed by linking the heavy and light chain fragments of the Fv region via an amino acid bridge, resulting in a single chain polypeptide.

30 Antibody fragments that recognize specific epitopes may be generated by known techniques. For example, such fragments include but are not limited to: the $F(ab')_2$ fragments, which can be produced by pepsin digestion of the antibody molecule and the Fab fragments, which can be generated by reducing the disulfide bridges of the $F(ab')_2$ fragments. Alternatively, Fab expression libraries may be constructed (Huse, et al., 1989, Science 246:1275-1281) to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity.

P 687 DK01

32

Immunoassays for gene products, conserved variants, or peptide fragments thereof will typically comprise incubating a sample, such as a biological fluid, a tissue extract, freshly harvested cells, or lysates of cells in the presence of a detectably labeled antibody capable of identifying gene product, conserved variants or peptide fragments thereof, and detecting the bound antibody by any of a number of techniques well-known in the art.

The biological sample may be brought in contact with and immobilized onto a solid phase support or carrier, such as nitrocellulose, that is capable of immobilizing cells, cell particles or soluble proteins. The support may then be washed with suitable buffers followed by treatment with the detectably labeled gene product specific antibody. The solid phase support may then be washed with the buffer a second time to remove unbound antibody. The amount of bound label on the solid support may then be detected by conventional means.

By "solid phase support or carrier" is intended any support capable of binding an antigen or an antibody. Well-known supports or carriers include glass, polystyrene, polypropylene, polyethylene, dextran, nylon, amylases, natural and modified celluloses, polyacrylamides, gabbros, and magnetite. The nature of the carrier can be either soluble to some extent or insoluble for the purposes of the present invention. The support material may have virtually any possible structural configuration so long as the coupled molecule is capable of binding to an antigen or antibody. Thus, the support configuration may be spherical, as in a bead, or cylindrical, as in the inside surface of a test tube, or the external surface of a rod. Alternatively, the surface may be flat such as a sheet, test strip, etc. Preferred supports include polystyrene beads. Those skilled in the art will know many other suitable carriers for binding antibody or antigen, or will be able to ascertain the same by use of routine experimentation.

One of the ways in which the RAI gene product-specific antibody can be detectably labeled is by linking the same to an enzyme, malate dehydrogenase, staphylococcal nuclease, delta-5-steroid isomerase, yeast alcohol dehydrogenase, α -glycerophosphate dehydrogenase, triose phosphate isomerase, horseradish peroxidase, alkaline phosphatase, asparaginase, glucose oxidase, β -galactosidase, ribonuclease, urease, catalase, glucose-6-phosphate dehydrogenase, glucoamylase and acetylcholinesterase. The detection can be accomplished by colorimetric methods

P 687 DK01

33

that employ a chromogenic substrate for the enzyme. Detection may also be accomplished by visual comparison of the extent of enzymatic reaction of a substrate in comparison with similarly prepared standards.

5 Detection may also be accomplished using any of a variety of other immunoassays. For example, by radioactively labeling the antibodies or antibody fragments, by labeling the antibody with a fluorescent compound. Among the most commonly used fluorescent labeling compounds are fluorescein isothiocyanate, rhodamine, phycoerythrin, phycocyanin, allophycocyanin, o-phthaldehyde and fluorescamine.

10

The antibody can also be detectably labeled using fluorescence emitting metals such as ^{152}Eu , or others of the lanthanide series or by coupling it to a chemiluminescent compound.

15

Diseases

Described herein are various applications of gene sequences, gene products, including peptide fragments and fusion proteins thereof, and of antibodies directed against gene products and peptide fragments thereof. Such applications include, for
20 example, prognostic and diagnostic evaluation of cancer and the identification of subjects with a predisposition to such disorders, as described above.

25

The method according to the invention may be used in relation to any cancer form, such as, but not limited to, skin carcinoma including malignant melanoma, breast cancer, lung cancer, colon cancer and other cancers in the gastro-intestinal tract, prostate cancer, lymphoma, leukemia, pancreas cancer, head and neck cancer, ovary cancer and other gynecological cancers. In particular the method is relevant for skin cancer, lung cancer, colon cancer and breast cancer, such as skin cancer and breast cancer, preferably wherein the skin cancer is basal cell carcinoma.

30

In particular, the method is relevant for early age cancer, such as early age breast cancer.

35

Gene nucleic acid sequences, described above, can be utilized for transferring recombinant nucleic acid sequences to cells and expressing said sequences in recipi-

P 687 DK01

34

ent cells. Such techniques can be used, for example, in marking cells or for the treatment of cancer. Such treatment can be in the form of gene replacement therapy. Specifically, one or more copies of a normal RAI gene or a portion of the RAI gene that directs the production of an RAI gene product exhibiting normal RAI gene function, may be inserted into the appropriate cells within a patient, using vectors that include, but are not limited to, adenovirus, adeno-associated virus, and retrovirus vectors, in addition to other particles that introduce DNA into cells, such as liposomes.

10 Examples

The examples relate to prediction from sequence polymorphisms in the region r to cancer. Blood was collected before (example 6) or after (examples 1 through 5) the persons acquired cancer. However, the sampling time is considered immaterial, as DNA in a polyclonal blood sample is not expected to change over time.

The particular sequence polymorphisms analysed in these examples are listed in Table 6, together with their sources of information and their definition as sequences.

P 687 DK01

35

Table 6. The markers used, their sources of information, and their currently estimated positions on chromosome 19, as well as their position in figure 2.

Name	Source of Identification	Position in sequence	GenBank Accession Number of sequence	Chromosome Position (Mbases)	Position in Figure 2
XRCC1 e10	Ref. 1	28152	L34079	59.420	1
CKM e8	rs#8188	20076	AC005781	61.361	2
XPD e23	Ref. 1	35931	L47234	61.479	3
XPD e10	Ref. 1	23591	L47234	61.491	4
XPD e6	Ref. 1	22541	L47234	62.4923	5
XPD i4	rs#1618536	19244	L47234	61.4924	6
RAI e6	rs#6966	8788	L47234	61.508	7
RAI i1	rs#1970764	875	L47234	61.514	8
ASE1 e1	rs#967591	232125	NT_011242	61.534	9
ERCC1 e4	Ref. 1	19007	M63796	61.547	10
FOSB e4	rs#1049698	34821	M89651	61.601	11
SLC1A5 e8	rs#1060043	60620	AC008622	62.946	12
GLTSCR1 e1	rs#1035938	20775	AC010519	63.986	13
LIG1 e6	rs#20580	111	L27710	65.460	14

rs numbers were derived from the NCBI's database dbSNP.

- 5 Ref 1: Shen, M.R., Jones, I.M., and Mohrenweiser, H. (1998) Nonconservative amino acid substitution variants exist at polymorphic frequency in DNA repair genes in healthy humans. *Cancer Res.*, 58: 604-8, 1998.

MATERIALS AND METHODS

- 10 *Study groups.* The groups of Caucasian Americans with and without BCC have been described previously (Athas et al, *Cancer Res.* 51:5786-5793, 1991; Wei et al, *Proc. Natl. Acad. Sci USA*, 90: 1614-8, 1994). Briefly, the study was a clinic based case control study at the Johns Hopkins Hospital, which serves multiple participating dermatologists in Maryland. Cases were histo-pathologically confirmed primary
- 15 BCCs and were diagnosed between 1987-1990. The controls were patients from the same physician practices and had a diagnosis of mild skin disorders. All participants were Caucasians living near Baltimore and were between 20 and 60 years of age.

P 687 DK01

36

- The controls were frequency matched to the cases by age and sex. Cases and controls with any other forms of cancer were excluded. In the questionnaire, the study subjects were asked if they had any blood relatives with skin cancer, and were asked to specify the type of cancer. Study subjects with relatives with basal cell carcinoma and squamous cell carcinoma and 'skin cancer' were included in the group of subjects with a family of skin cancer. Subjects with relatives with melanoma were not included. At the clinic visit the subjects gave informed consent, were examined by dermatologists, completed a structured questionnaire and provided blood. DNAs from available frozen lymphocytes were purified using Puregene (Gentra Systems) and were genotyped. Initially, 71 cases and 118 controls were included in this study. However, the number of persons varied between analyses, as the supply of DNAs gradually was depleted. In case of the SNP RAI 11 only 133 persons could be genotyped reliably.
- The groups of 20 psoriatic Danes with and 20 psoriatic Danes without BCC have been described previously (Dybdahl et al, Cancer Epidemiol. Biomarkers Prev., 8:77-81, 1999). Briefly, BCC subjects were identified from a population-based cohort of persons treated by Danish dermatologists in the year 1995, and fulfilled the following criteria (a) age in 1995 < 50 years; and (b) clinically verified diagnosis of psoriasis. The diagnosis of BCC was clinically and histologically confirmed. The controls consisting of psoriasis cases without BCC was selected from among patients treated in the year 1992-1995 for psoriasis by dermatologists who participated in the national cohort study 1995. The controls were matched by age and sex. The patients with psoriasis and BCC differed from the national cohort of BCC in that the average of first BCC was 38 year against 56 year in the cohort. A number of cases had had multiple BCCs. There was a tendency that cases had been treated for a longer time than the controls, and also that the treatments were more intense. This was to be expected as treatment of psoriasis involves a number of carcinogenic treatment modalities. DNAs from available frozen lymphocytes were purified using Puregene (Gentra Systems) and were genotyped.

Primers and probes. Table 7 includes the polymorphisms typed on Lightcycler™, the primers used for the PCR reaction and the probes used for detection and typing of the PCR products. Table 8 lists the polymorphisms typed by conventional PCR-RFLP, and the primers and restriction enzymes used. Table 9 lists the polymor-

P 887 DK01

37

phisms typed by SNaPshot technology and the primers used. Table 10 lists the polymorphisms analyzed on a Taqman, and the primers and probes used. Hobolth DNA, Hillerød, Denmark or DNA Technology, Aarhus, Denmark, synthesized the primers in tables 7, 8, and 9. TIB Mol-Biol, Berlin Germany synthesized the Lightcycler probes. TAG-Copenhagen ApS (Tagc.com, Copenhagen, Denmark) synthesized the primers, and Applied Biosystem synthesized the fluorescent Taqman probes in table 10.

Table 7. Design of primers and fluorogenic probes for LightCycler

ASE1 e1

Forward primer: 5'-GGTTTTCTGCTCTGCACACG

Reverse primer: 5'-CCTTTCTCCTTCCACCAACG

Anchor probe: 5'-TCTGCAACCTGGTGCGAGCAGC-Fluorescein

Sensor probe: 5'-LCRed640-CGGGCTACAGGGTTACCTGAG-p

CKM e8

Forward primer: 5'-TTGAAACTGGAAGCTCTGAGAAGG

Reverse primer: 5'-TGGTGGATGGTGTGAAGCA

Anchor probe: 5'-LC Red 640-

CCTTTCTCCAAGTCTTCTCCATTTCACG-p

Sensor probe: 5'-GGGGATCATGTCGTCAATGGACT-Fluorescein

ERCC1 e4

Forward primer: 5'-AGGACCACAGGACACGCAGA-3'

Reverse primer: 5'-CATAGAACAGTCCAGAACAC-3'

Anchor probe: 5'-LCRed640-TGGCGACGTAATTCCCGACTATGTGCTG p-3'

Sensor probe: 5'-CGCAACGTGCCCTGGGAAT-Fluorescein

FOSB e4

Forward primer: 5'-AGGCTCAACAAGGAAAAATGC

Reverse primer: 5'-GCTAGACAGTCAAGGAGGGACG

Anchor probe: 5'-LCRed 640-AAAGGGTGGGTGTGGGAGACATTGG-p

Sensor probe: 5'-AAACCAACCTAGGCACCCCAAA-Fluorescein

GLTSCR1 e1

Forward primer: 5'-CGACGAACTTCTCTGAAGCGAA

Reverse primer: 5'-AGCGACACGGGCATCTGG

Anchor probe: 5'-ATGAGCGTCCACCTCCTGAACC-fluorescein

P 687 DK01

38

Sensor probe: 5'-LCRed 640-AGGCAGCAGCATCGTCATCCCC-p

LIG1 e6

Forward primer: 5'-ATGCCCTGTAGGTTCAATGG

Reverse primer: 5'-TGGAGGTCTTTAGGGGCTTG

Anchor probe: 5'-GGCTGGTCCCCGTCTTCTCCTTCC-Fluorescein

Sensor probe: 5'-LC Red 640-TCTCTGTTGCCACTTCAGCCTC-p

RAI1

Forward primer: 5'-TGGCTAACACGGTGAAACC

Reverse primer: 5'-GGAATCCAAAGATTCTATGATGG

Anchor probe: 5'-GGGAGGCGGAGCTTGCACTGA-Fluorescein

Sensor probe: 5'-LCRed 640-CTGAGATCGCACCCTGCAC-p

SLC1A5 e8

Forward primer: 5'-CAGTGTCCAAAGAGCACC

Reverse primer: 5'-CTACCCCTTTAGCGACC

Anchor probe: 5'-LCRed 640-TCCTGCCCCCAGAGCGTCACC-p

Sensor probe: 5'-GTACGGTCCACATAATTTTGGAGGA-Fluorescein

XPD e10

Forward primer: 5'-GATCAAAGAGACAGACGAGC

Reverse primer: 5'-GAAGCCCAGGAAATGC

Anchor probe: 5'-GGACGCCCACCTGGCCAACC-Fluorescein

Sensor probe: 5'-LCRed640-CGTGCTGCCCAACGAAGTG-p

Table 8. Primers and restriction enzymes used for typing of SNPs using PCR-RFLP

Gene exon	Primers	Enzyme	Digested Fragments
XRCC1 exon10	TTGTGCTTTCTCTGTGTCCA TATCAGAAAAGGCTGGAGGA	MspI	240, 375bp (A) 615bp (G)
ERCC1 exon4	AGGACCACAGGACACGCAGA CATAGAACAGTCCAGAACAC	BsrDI	167, 368bp (A); 525bp (G)
XPD exon8	1.set CACACCTGGCTCATTTTTGTAT TCATCCAGGTTGTAGATGCCA	T7II	
	2.set TGGAGTGCTATGGCACGATCTCT CCATGGGCATCAAATTCCTGGGA	T7II	56, 114, 482 bp (A); 66, 596 bp (C)

P 687 DK01

39

XPD exon23 1.set	GTCTGCCCTCAGCAAAGAGAA		
	TTCTCCTGCGATTAAAGGCTGT		
	ATCCTGTCCCTACTGGCCATTC	<i>PstI</i>	66, 100, 158 (C);
	TGTGAACGTGACAGTGAGAAAT		100, 224 (A)

Table 9. Design of primers and SNaPshot primers for SNaPshot typing on sequenator.

XRCC1 exon7

Forward primer: 5'-GTCCCATAGATAGGAGTGAAAG

Reverse primer: 5'-CCCTAGGACACAGGAGCACA

SNaPshot primer: 5'-TGCATAGCTAGGTCCTGC

XRCC1 exon17

Forward primer: 5'-GCCAAGCAGAAGAGACAAA

Reverse primer: 5'-GAGTGGCTGGGGAGTAGGA

SNaPshot primer:

5'-AACTGACRAAACTAGCTCTATGGGGTGGTGCCGCA

RAI exon6

Forward primer: 5'-CCTACCACCATCATCACATCC

Reverse primer: 5'-GCCTTGCCAAAATCATAACC

SNaPshot primer: 5'-CCTCTCCCAATTAAGTGCCTTCACACAGC

XPD intron4

Forward primer: 5'-CGCAAAAACCTTGTTATTCACC

Reverse primer: 5'-CCCATTTTATCATCAGCAACC

SNaPshot primer: 5'-CTGGCTCTGAACTTACTAGCCC

Table 10. Design of primers and probes for Tagman.

XRCC1 exon10

Forward primer: 5'-GCT GGA CTG TCA CCG CAT G

Reverse Primer: 5'-GGA GCA GGG TTG GCG TG

Probe (A): 5'-Fam- TGC CCT CCC AGA GGT AAG GCC T -Tamra

Probe (G): 5'-Vic - CCC TCC CGG AGG TAA GGC CTC -Tamra

P 687 DK01

40

Determination of polymorphisms by Lightcycler. Genotypes of the American persons for polymorphisms in ASE-1e1, CKMe8, ERCC1e4, FOSBe4, GLTSCR1e1, LIG1e6, RAI1, SLC1A5e8 and XPDe10 and of the Danish persons for polymorphisms ASE-1e1, CKMe8, FOSBe4, LIG1e6 and SLC1A5e8 were detected using LightCycler™ (Roche Molecular Biochemicals, Mannheim, Germany). PCR was performed by rapid-cycling in a reaction volume of 20 µl with 0.5 µM of each primer, 0.045 µM of anchor and sensor probe, 3.5 mM MgCl₂, approximately 7 - 25 ng genomic DNA, and 2 µl LightCycler DNA Master Hybridization probe buffer (Roche Molecular Biochemicals, Cat. No 2158 825). This buffer contains Taq DNA polymerase, dNTP mix, and 10 mM MgCl₂. In some cases the reaction mixture also contained 5% DMSO. The temperature cycling consisted of denaturation at 95°C for 2 sec, followed by 46 cycles consisting of 2 sec at 95°C, 10 sec at 57°C, and 30 sec at 72°C. The last annealing period at 72°C was extended to 120 sec. The melting profile was determined by a temperature ramp from 50°C to 95°C with a rate of 0.1 degree/sec. For RAI12 the melting profile was run 3 times, and the last curve was used.

PCR-RFLP analyses. Genotypes of the American persons for polymorphisms in XPDe6 and XPDe23 and of Danish psoriatics for polymorphisms in XRCC1e10, ERCC1e4, XPDe6, and XPDe23 were detected using PCR-RFLP technique (Shen et al see above; Dybdahl et al, see above; Vogel et al, Cancer Epidemiol. Biomarkers Prev., 8:77-81 (2001)). The reactions were performed as reported (Shen et al, see above; Dybdahl et al, see above; Vogel et al, Cancer Epidemiol. Biomarkers Prev., 8:77-81 (2001)).

Determination of polymorphisms by SNaPshot technique on sequenator. The polymorphisms in RAIe6, XPDI4, XRCC1e7, and XRCC1e17 in the American persons were typed simultaneously on an ABI Prism 310 sequenator (Applied Biosystems, Foster City, CA, USA) using SNaPshot technique (Lindblad-Toh et al, Nature Genetics, 24: 381-6, 2000.). The PCR reaction consisted of 1 µl of purified genomic DNA, 1 pmole of each primer (DNA Technology, Aarhus Denmark), 12.5 nmole of each dNTP (Bioline, London, UK), 100 nmole MgCl₂ (Bioline), 0.15 µl BIOTAQ™ DNA Polymerase (Bioline) in a total volume of 20 µl of water. The program consisted of 4 min at 96°C, followed by 25 cycles of 96°C for 30 sec, 60°C for 30 sec, and 72°C for 60 sec. The last cycle was followed by 72°C for 6 min. The primers and dNTPs were removed in reactions containing 2 U Shrimp Alkaline Phosphatase

P 687 DK01

41

(SAP) (Roche), 2 U Exonuclease I (Biolabs, Denmark), and 9 μ l PCR reaction in a total volume of 14 μ l water. The reactions were incubated at 37°C for 60 min and 72°C for 15 min. The SNaPshot reactions contained 1 μ l of SNaPshot Ready Reaction Mix (Applied Biosystems), 0.5 μ l of each SNaPshot primers (XRCCe7-ss1; 4pmol/ μ l, XPDi5-cp1; 0.5pmol/ μ l, RAla7-cp1; 1pmol/ μ l; XRCCe17-ss1; 2pmol/ μ l), 2 μ l of the purified PCR product, and 1.5 μ l of buffer (200 mM Tris-HCl, 5 mM MgCl₂, pH 9.0). The reactions were cycled 25 times: 96°C for 10s, 50°C for 5s, and 60°C for 30s. The primers and dNTPs were removed in a reaction containing 1 U SAP, 0.8 μ l 10xSAP buffer, and 5 μ l SNaPshot reaction in a total volume of 8 μ l of water. Two μ l purified product was added to 10 μ l of concentrated deionized formamide (Amresco, Ohio, USA), incubated for 5 min at 95°C, and analyzed on the sequenator. The two markers in XRCC1, in exon 7 and exon 17, could not be reliably scored and thus were excluded from further consideration.

Determination of polymorphisms by real-time PCR using Taqman probes. The polymorphism in XRCC1e10 in the American persons was analysed using the ABI Prism 7700 sequence detection system (Applied Biosystems, Foster City, Ca, USA). PCR Primers and Taqman probes were designed using Primer Express v 1.0 (Applied Biosystems). The reactions were performed in MicroAmp optical tubes sealed with MicroAmp optical caps (Applied Biosystems) containing a 10 μ l reaction volume: 1x Taqman buffer A, 2.5mM MgCl₂, 200 μ M each of dATP dCTP, dGTP, 400 μ M dUTP, 800nM each primer, 200nm each probe, 0.01U/ μ L AmpErase UNG, 0.025 U/ μ L AmpliTaq Gold Polymerase. Thermal cycler conditions were: Tubes were incubated at 50°C for 2 min followed 10 min at 95°C. The incubation was succeeded by 45 cycles of 95°C for 15 sec and 64°C for 1 min.

Example 1

DNA from humans from the American cohort of patients with basal cell carcinoma and controls, described in Materials and Methods, was typed with respect to a number of sequence polymorphisms located in and around the claimed region r. The resulting statistical p-values for association of occurrence of the individual sequence polymorphisms with the status of patients are depicted in Figure 2. Also depicted are the calculated odds ratios for association of sequence polymorphism and disease. For the calculation of the odds ratios the heterozygote genotypes were combined

P 687 DK01

42

with the lesser group of homozygotes, and the ordering of the groups was chosen such that the odds ratio became more than or equal to 1. The results show that the sequence polymorphism RAl1 is strongly associated with disease in this cohort ($p = 0.004$). Bonferroni correction for the number of tests made indicates that a result less than 0.007 must be considered significant at a level of 0.05. Thus, even after correction for multiplicity of testing this result is significant.

The numbers next to the points in the curves are merely a help to identify the single sequence polymorphisms:

1, Xr1e10; 2, CKMe8; 3, XPDe23; 4, XPDe10; 5, XPDe6; 6, XPDI4; 7, RAl6; 8, RAl1; 9, ASE-1e3; 10, ERCC1e4; 11, FOSBe4; 12, SLC1A5e8; 13, GLTSCR1e1; 14, LIG1e6.

Example 2

Those persons in example 1, who got basal cell carcinoma before the age of 50 years, were selected, and the results from analysis of RAl1 were compared the status of the patients. There was a strong relationship between the occurrence of the individual genotypes of the sequence polymorphism and the status of the patients (Table 7; Odds ratio = 12.3; $p(\chi^2) = 0.00014$).

Table 7. Occurrences of genotype for the sequence polymorphism RAl1 in American with Basal cell carcinoma occurring before 50 years of age and in controls.

RAl1 genotypes	Number of cases before 50 years of age	Number of controls
AA	31	44
AG	2	32
GG	0	5

Example 3

The data of Example 2 were combined with results of genotyping the neighbouring sequence polymorphism RAl6. There was a very strong association between the

P 687 DK01

43

combined genotypes of RAI1 and RAI6 and the status of the patients. Thus, almost all American cases occurring before the age of 50 yrs were homozygote for RAI 11^A RAI e6^A, while only approximately half of the controls were so (Table 8, Odds ratio = 12.8; $p(\chi^2) = 0.00006$).

5

Table 8. Combined occurrences of different genotypes for the sequence polymorphisms RAI2 and RAI7 in American cases occurring before 50 years of age and in controls.

		RAI1		
RAI6		AA	AG	GG
BCC cases	AA	30	0	0
	AT	0	2	0
	TT	0	0	0
Controls	AA	42	10	1
	AT	2	21	0
	TT	1	0	2

10

Example 4

The data of Example 2 were combined with results of genotyping the sequence polymorphism GLTSCR1e1 located outside the claimed region r. There was a very strong association between the combined genotypes of RAI1 and GLTSCR1e1 and the status of the patients. It was obvious to define "risk-genotypes" as having two As in RAI1 and at least one C in GLTSCR1e1. This corresponds to the assumptions that RAI1^A is recessive, and GLTSCR1e1^C is dominant. If one does so, one finds that 26 out of 25 cases have a "risk-genotype", while only 28 out of 62 controls have one (Table 9; Odds ratio > 30; $p(\chi^2) = 0.000002$).

15

20

P 887 DK01

44

Table 9. Combined occurrences of genotypes for the sequence polymorphisms RAI1 and GLTSCR1e1 in American cases of basal cell carcinoma occurring before 50 years of age and in controls.

		RAI1		
GLTSCR1e1		AA	AG	GG
BCC cases	CC	17	0	0
	CT	8	0	0
	TT	0	0	0
Controls	CC	15	18	3
	CT	13	7	0
	TT	3	3	0

5

Example 5

DNA from humans from the cohort of Danish psoriatics with basal cell carcinoma and controls, described in Materials and Methods, was typed with respect to a number of sequence polymorphisms located in and around the claimed region *r*. The resulting statistical *p*-values for association of occurrence of the individual sequence polymorphisms with the status of patients are depicted in Figure 3. The results show that the sequence polymorphism ERCC1e4 is strongly associated with disease in this cohort ($p = 0.01$).

15

Example 6

Blood samples were collected from a large number of Danish citizens and frozen. After a number of years those women, who got breast cancer in the intervening period, were identified, as well as a set of matching controls. DNAs were purified from the blood samples of these persons and a number of polymorphisms, namely RAI1, ASE-1e3 and ERCC1e4, in the region of interest were typed. The polymorphisms were subsequently combined such that the high-risk group was homozygous for the high-risk alleles of all three polymorphisms: RAI1^{AA}ASE-1e3^{GG}ERCC1e4^{GG}. All other genotypes were combined into the low-risk group (Table 10; OR = 1.59; $p(\chi^2) = 0.004$).

20

25

P 687 DK01

45

Table 10. Occurrence of a combined "high-risk" genotype RAI1^{AA}ASE-1e3^{GG}ERCC1e4^{GG} as opposed to all other combinations of genotypes for the sequence polymorphisms RAI1, ASE-e3 and ERCC1e4 in Danish cases of breast cancer and controls.

5

	High -risk	Low-risk
Cases	120	85
Controls	277	312

10

The DNAs in these examples were purified from available frozen lymphocytes using Puregene (Gentra Systems). A variety of other ways of purifying DNA is available to the expert and would also be expected to lead to the wanted results.

15

Analysis of sequence polymorphisms can be performed with a variety of techniques, some of which have been used in the examples of this application. Most often a number of techniques can produce the wanted result.

Similarly, the choice of primers and probes in a particular assay is to some extent free and other primers and probes might well produce similar results.

20

Finally, it is to be expected that assays for other sequence polymorphisms in the region of interest may produce roughly similar results. Our particular choice of sequence polymorphisms and assays used in the examples are thus not intended to limit our claims. Thus, at present about 30 SNPs within the region *r* are listed in NCBI's database dbSNP including rs#2070830, rs#2017104, rs#2017154 and rs#2377328, all within or very close to RAI. Other forms of polymorphisms such as the tandem repeat polymorphisms D19S543 and D19S393 are also known to occur in the region and can probably serve as markers in the present invention. Moreover, it is very likely that the region contains a number of as yet undiscovered polymorphisms. For instance, the sequence of the 5' half of RAI and its upstream promoter region is currently only a draft version and new polymorphisms of potential use for this invention are likely to be uncovered as more sequence reads of this segment are produced.

30

P 887 DK01

Modtaget

48

- 7 OKT. 2002

Sequence of the r region of chromosome 19

PVS

The following depicts the region r stretching from the beginning of, but not including the XPD gene, to approximately the end of ERCC1, and includes the genes RAI, LOC182978, and ASE-1. More specifically r is bounded by and includes the following two sequences: AGAACCCCCG CCCCTCCACC TCGTCTCAAA and TCCCTCCCCA GAGACTGCAC CAGCGCAGCC, and is defined by SEQ ID NO: 1 herein below:

10 AGAACCCCCGCCCCTCCACCTCGTCTCAAAAAAAAAAAAAAAAAATCGTCTCAGTAGCGA-
ATAGTCTAACGGAGAATGACAGGGAAATTGGTGATCCTTTCTGGGCCCAAGAGTTA-
GAAATGGCTTTGCAGGCCGGCGCGGTGGCTCAAGCCTGTAATCCCAGCACTTTGG-
GAGGCTGAGGCAGGTGGATCACCTGAGGTCGGGAGTTCAAGACCAGCCTGACCAA-
CATGGAGAAAACCTGTCTCTACTAAAGATACAAAATTAGCCGGGCGTGTGGCAAATG-
15 CTTGTAATCCCAGCTACTCGGGAGGCTGAAGCAGGAGAATTGCTTGAACCTGGGAGG-
CAGAGGTTGCAGTGAGCAGAGATGGCGCCGTCCGACTCTAGCCTGGGCAACAAAAG-
CGAAACTCCATTTCAAATATTAATAATAAATAAATAAATAAATAAATAAATAAATAA-
TTGTTTGTCTTCAACAAATAGCTATGTGGCATCTACCATGTGTCTGATCCTGTGCT-
GGCCCCCTGGGAACAGAAAGGTGACCATGACAGCCTCAGCACCTGCCCTCAAAGAACA-
20 GATTTTTTCTTGGAGACAGGGTCTTTCTGTGTCGCCAAGGCTGGAGTGCACTGGCA-
CAGTCACAGCTCACTGCAGCCTCCACCTCTTGGGCTCAAGCGATCCTCCACCTCAG-
CTTCCAGAGTAGCTGGGACCACAGGTGTGCACCACCAAGCCAGCTAAGTTTTATTTT-
TTAAATTTTTTAGAGACGAGGTCTCACCACGTTGCCCAGGCTGGTTAAACTCGCAG-
GTTCAAGTGATCCTCTCCCCTCAGCCTTTCAAATTGTTGGGATTACAGGGGTGAGG-
25 CACCAGGCCTGGCCTCAAAGAACAGATATTAATATACAAATGAATATATGATTACAGC-
CTGGAGTGGTGGCTCGTGCTGTGTTCCAACTTTGGAAGGCCAAGGCCAGTA-
CATTGCTTGAGCTCAGGAGCTAGACCAAGCCTGGGCAACATGGTGAAAACCCGTC-
TCTACAAAAAATGCAAAATTAGCTGGGCGTGGTGGCGTGCACCTGTAGTCCCA-
TACTCAGGAGGCTGAGGTGGGAGAATCACCTGGGCCTGGGAGGCAGAGGTTGCAAT-
30 GGGCAGTGATTGTGCCACTGCACTCCAGCCTGGGCAACAGGAGTGAAAACCTATCT-
CAAATGT-
CAAAAGTGCTGTGAAGGAAAACCTCAAGTCACCATAAAGATTGATTATGGGCTGGTG-
CAGTGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCCAAGGCAGATGGAT-
CACGAGGTGAGGAGTTCAAGACCAGCCTGGTCAACATGGTGAAACCCTATCTCTAC-
35 TAAAAAAAAAAAAAAAAAAAAAAAAAGCCAGGCATAGTGGCATGCATCTGTAATCCCATC-
TACTCGGGAGGCTAAAGCAGGAGAATTGCTTGAACCCAGGAGGCAGAAAGTGAGCCAA-
GATCACGCCACTGCACTCCAGCCTGCGTGACAGAGCAAGACTCCGTCCCAGAAAAA-
GAAAAAAAAAAAAAGACTTATTATGACAGGATGTCTACTGTCAACTGTGGGGTGTGAGT-
GTTGGCCAAGTGATCAGAGAAGGCTTCGTGGAAGAAGCGAGGTTTGAGTAGAGCCA-
40 GAAAAATAATTAGAAGAGATCAACCAGCAAGAGGGGATGGATGAGAGAAGTGAGAAAG-
GTGTTCCAGGGAGAGAGACCATCATACAAAAAGCTCTAGGCCAGAAAGAAAGCT-
GAGGCCTGTGAGTGCTGAAAGGAAGCCTGTGGGGGTGGAGCTCTGAGTTGAGCA-
CAGGGAGCAGAGAAAGGGCAGCTGGAGGGGAAGGCAGGGGCAGATCGAAATCTCTT-
TTTTAAATTAATTAATTTCTTAATTTATTTTATTTTATTTTATTTTATTTTATTTT-
45 CAGACTGGAGTACAGTGGCACAATCTCAGCGCACCGCAACCTCTGCCACCCAGGCT-
CAAGCAATTTCTGTGCCCTCAGCCTCCCTAGTAGCTGGGATTACAGGTGCGCACCAC-
TACTGCCACAGCTAATTTTATACTTTTAGTAGAAACGGGGTTTCACTATGTTGGCCAGG-
CTGGCCTCAAACCTGACCTCAAAAGATCCACCCACTTCAGCCTCCCAAAGTGCTG-
GGATTACAGGTGTGAGCCACCCTTCCCGGCTGTATTTTGGAGACAGAGTCTTGCTCT-
50 GTCCAGCCTGGAGTATGGTGGTGTGAATTTGGCTCATTGCCACCTTGACCTCCAG-
GGCTCAAGTGATCCTCCACCTCAGCCTCCTGAGTAGCTGGGACTGCGGGTACACGA-
CACCACGCCTGGTTAATTTTTTTTAAATTTTTGTAGAGACGAGGGTATCTCACTATGTT-

P 687 DK01

47

GTCCAGGCTGGTTGAACTCCTGAGCTCAAGCAATTCTCCACCTCAGCCTCC-
CAAAGTGGTGGGATTACAGACGTGAGCCACTGTGCCCGGCTTAATTTATTACATAA-
ATTTTTTATGTTACTTTTCTATCTCCTACAGGAAGAAAATATATTTTGTATTGACAG-
GGTCTCGCTATGTTGCCAGGCTGGTATTGGGCTCAAGCCATCCTGTTCCCTCAGCC-
5 TCCCAAAGTACTGGGATTACAAGCGTGAGCCTCTGCATCCAGCCCAGATCCAAAATCT-
TACTGTCACCTACAGAGTCTCTGTAAGTACTGCTTACTGCTCATCATCCCCATACCAAC-
CCACCTTACTGCTCTGATCTCCTCCTCTCTCTCCCCAGCTCATTTTGTTCAGCTATG-
CTGGTCTCCTTGTCTCTCTAAAACATAACAAGCACATCCCATCTCAGGGCCTTTG-
CACCAGCTATTTTGTCTGCCTGGAATGCTGTTTCCCCTGATAGCCATGTGGCTGACA-
10 CACTCACCTCCCTCAGCTCTTTGCTCAATTGTCAACTTCTCGGCCCGGCATGGTGGCT-
CACACCTGTAATCCTACCACTTTGGGAGGCTGAGGTGGGCAGATCACCTGAGATCAG-
GAGTTCGAGACCAGCCTGGCCAAGATGGTGAAATCCCGTCTCTACTAAAAATACAAAA-
ATTGGCAAAGCATGGTAGCACATACCAGTAATCCTAGCTACCCGGGAGGCTGAGG-
CAGGAGAATTGCTGGAACCCGGGAGGCTGAGGCTGCAGTGAGCCAAGATCATGC-
15 CACTGTACTCCAGCCTGGGTGACAAAGCAAGACTCTGTCTCAAAAAAAAAAAGTCTC-
CTTCTCAATGAGGGCTTCTGACCACCAAAATTAATCTACCTCCTAGACACACACACA-
CACGCACGCACGCACGCACACACACACGCACGCACGCACACACACACACACACA-
CACACTATATCCCCCTTTCCCTGCTTTATTGTTCTTGAGAGCTCATTTAACCATGTGA-
CATGCTGAATATTTTACTTATTTATTTTGTAGAAAGCTCCTGGCTGGGCGGGGGC-
20 TCACGCCTGTAATCCAGCACTTTGGGAGGCTGGAACAGGTGGATCATGTGAGGT-
CAGGAGTTCCAGACCAGCCTGACCAACACGGTGAAACCTCATCTCTATTAATAATG-
CAAAAATTAGCTGGGTGTGGTGTGCTGCCTGTAATCCCACTACTCAGAAGGCT-
GAAGCAGGAGAATCGCTTGAACCTGGGAGGCAGAGGTTAACGCTGAGCCGAGATCG-
CGCCATTGCACTCCAGCCTGGGCAACAAGAGTGAAACTCTGTCTCGAAAAAA-
25 CAAAAGTCAAGTCCATGGCAGGAGTGATGGCTCACGCCTATAATCCAGCACTTTGT-
GAGGCCGAGGCGGGCGGATCACTTGAAGTCAAGGAGTTGGAGACCAGCCTGGCCAA-
CATGGTGAAACCTCATCTCTACTAAAAATACAAAAATTAGCCGGGCGTGGTGACACAT-
GTCTGTAGTCCAGCTACTTGGGAGGCTGAGGCTGGAGAATGGCTTGAACCTGG-
GAGGTAGAGGTTGCAGTAAGCCAAGATCGCGCCATTGCTCTCCATCCTGGGCAACA-
30 GACTCCGTCTCAGAAAGGAAGAAAGGAAGGAAGAGAGAAAGAGAGAAAGAGACAGA-
GAGAGAGAGAGAAAGGAGAAAGAGAGAAAGGATGGAAGGACCCTGACAAGCACT-
GTTGCATAAAAGTTTCTTTTCTCTCTTTTTTTTTTTTTTTTTTTTTTTTGGAGAGGGTC-
TCACTTCTGTTGCTCCAGCTGAAGTGCAGTGGTGAGAACATGGCTCAGTGCAGCCT-
CAACTTCCAGGCTTAAGTGATCCTGCCACCTCAGCCTCCTGAGTAGCTGGGACTG-
35 TAGGTGTGCACCACCGTGCCATAGCTAATTTTTTTGTATTTTATAGTAGACATGGTTCCG-
CCACGTTGCCAGGCTGGTCTTGAACCTCCTGGGCTTAAGGGATCTGCCGGCCATGGC-
CTCCCAAAGTGCTGGGATTACCAGCGTGAGCCACTGTACCCAGCCTGAGTATAGGTT-
TCTGATAAATTTTAGGATCATATTGTTGGACTGGGTAAGAATTTCCAGAACTCTAAT-
GAAGAACTGACTGGTTTATATTTTATTTTATTTTATTTTATTTTATTTTATTTT-
40 TCACTCTTGTGGCCAAGCTGGATTGCAGTGGCACGATCTTGGCTCACCACAACCTC-
CGCCTCCCGGTTTCAAGTGATTCTCCTGCCTCAGCCTCCCCAGGAGCTGGGATTA-
CAGGCACCCACCACCATGCTCGGCTATTTTTTTTTTTATTTTTTATTTTTAGTAGA-
GACGGGTTTTACCATGTTGGCCAGGCTGGTCTCGAACTCCTGACCTCAGGTGATC-
CACCTGCCTTGGCCTCCCAAAGCGCTGGGATTACAGGCATGAGCCACTGTGCAAGGC-
45 CTAGGCTGGTTTATAAAATTGCTAAACCAAGCAGAACATGAATTAATACCAAGGAA-
ATACTCTCCTAGATTGTCATGTTACATCAGCCAATACTAAAATTGTCAAGATACACAAT-
TTGAATGAACTCCATGGTCCAAGTCGAATTATCTATGATATTACCCATCTAATAAACAG-
CACTATGTCCCTTAATGGGAGAAAAAGTTGGAGAAATTAAGAGAATATCAATCCAAT-
GTTGGTTGGGTGCAGTGAATCATGTCTATATTTCCAGCACTTTGGGAGGCCAAGG-
50 CAGGAGGATCACTTGAGCCCAGGAATTCAAGGCCAGCCTCGGCAACACGGTGAGATC-
CTGTCTCTACGGAAAAATTAATAAAAAAAAAAAGAGAGAGATTAGTGGGATGTGGTGCC-
TATAGTCCCAGCTACTTGGGAGGCTGAGGCGGGAGGATCATTTAAGCCTGGGACGTT-
GAGGTTGCAGTGAACCATGAGTGAGACTCATCTCAAAAAAAAAAAAAAAAAAATGGCGAT-
CACTAGAGGAAAAAAAAAACTAAAGTGGGGTTTGGCGGTAGTGGGAGGGCCCTTCCTG-
55 CTAGGTTGCACATGATCTCCAGGAGGCTCCACGGGAGAATCATTTCTTGTCTTTT-
TCAGTTTCTAGAGCCAAATTCTTTGCAATACCTTGCAATCCTTGGCTCGGAACCCCTTCC-
CTAACCTTCAAAGCTGGCAGCTAGCCTCTGGCTCAAGTGTACATGGCCTGTCTCT-
GTCTTCTATCCAATCTTCTCTTATAAGAACATTGGAGCCAGGCATGGTGGCTGACG-

P 687 DK01

48

CCTGTAATCCCAGCACTTTGGGAGACCGAGGCAGGCGGATCACAAGGTCAGGAGT-
TCGAGACCAGCCTGGCCAACACAGTGAAACCCCGTCTCTACTAAAAAATA-
CAAAAAAGTAGCCGGGCATGGTGGCAGGTGCCTGTAATCCCAGCTACTTGAGAGGCT-
5 GAGGCAGGAGAATCGCTTGAACCTGGGAGGCAGAGCTTGCACTGAGCCGAGATAGT-
GCCAATGCAGTCCGGCCCTGGGCGAAACAGCGAGACTCCGTCGCAAAAAAATAA-
ATAATAATAATAATAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAA-
ATAAAAAATTATTTGAGACAAAGTCTATTCTGTGGCAGAGGCTGGAATGCAGTGGCGT-
GATCACAGCTTACTGCAGCTTCTACCTCCTGAGCTCAAGCGATCCTTCCACCTTGGCT-
10 TCCTGAGTAGCTGGGACCTCAGGTGTACATTACCACGCTCAGCTAATTATTTATT-
TATTATATTTTGTGACGGAGTTTCGCTCTTGTGGCCGGGCTGGAGTGCAATGGTGC-
TATCTCAGCTCACTGCAACCTCTGCCTCCTGGATTCCAGTGATTCTCCTGTCTCAGCT-
TCCTGAGTAGCTGGGATTACAGGTACATGCCATCACGCCAGCTAATTTTGTATTTT-
TAGTAGAGACGGGGTTTCATCATATTGGTCAGGCTGGTCTCGAACTCCTGACCTCAG-
15 GTGATCCACCTGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGGCACACG-
CCCGGCAATTTTTTTTTCTTTTTTTTTTTCAGACAGAGTCTTGCTCTGTCAACCAGGC-
TGGAGTGCAGTAGCGTGATCTCGGTTTACTGCAACCTCCATCTCCCGGGTTCAAG-
CGATTCTCCTTTCTCAGCCTCCCAAGTAGCTGGGACTACAGGTGCACACCACCACGG-
CGGGCTAATTTTTGTATTTTGTAGTACACACAGGTTTCCACATATTGGTCAGATGGT-
20 TCAAACTCCTGACCTCAGGTGATCCATCTGCCTCAGCCTCCCAAATTTGGGATTA-
CAAGCGTGAGCCACACACCTGGCTTAATTTTTTATTTTGTATCGACACAGGGTCTCCC-
TATGTTGTCCAAGCTGGCAGAGATTTTGTGTTGTTGTTGAGAGGGAAATTTGCTCTT-
GTAGCCCAGGCTGGAGTACAATGGTGCAATCTTGGCTCACCACAACCTCCGCCTCC-
CGGGTTTAACAGATTCTCCTGCCTCAGCCTCCCAAGTAGCTGGAACACAGGCACC-
25 TACCACCACACCAGGCTAATTTTTGTGCTTTTGTAGTAGAGATGAGGTTTACCATGTT-
GGCCAGGCTGGTCTTAACTCCTGGCCTCCAGTGATCCACCCGCTTGACCTCC-
CAAAGTGCTGAAATTACAGGCGTGAGCACCAGGCGCCTGGCCTCTCAACCTACAATTT-
CAACACCCCAAGGAAACAGCCACCATGAGTGAGAACCAGCAGACACAACAATA-
TAGGATTAGCTGCCTCCAACTTCAGGTGATAGATTATCAGGCATGTACTTGAAAC-
30 TAAAGGACACAAAAGAAGATCCGAAATATAAAATAAAGGATTGGACTTGTGTGAAAA-
GAATCCCTTAGAAAGGGCTACTTTCAGGCTGGCCATGGTGGCTAATGGCCTGTAATC-
CCAGCACTTTGGAAGGCCGAGGTGTGTGGATCACCTGAGGTCAAGAGTTCAAGAC-
CAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTGAAAATACAAAAATTAGCCAGGT-
GGGGTGGCAGATGCCTGTAATCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCGC-
35 TTGAACTCAGGAGGCAGAGGTTGCAGTGAGCTGAGATTGCGCTATCGTGCCCCAGCC-
TGGGCACTAGAGTGAGATCAAAAAAAGAAAAAAGAAAGAAAGAAAGGGC-
TACTTTCAGACTGCCTTGCCAAAAATCATAACCACAATGATGAGCATGTATTGAGT-
CAAAACAGAATCAAAAGAGAAGAAAGTCAATTTCTGTGCAAACTACTTTTATTATAAG-
GAAAGTTTCTCTATTTTGTATAAACATTAACCAGTGCTGTGTGAAGGCATTAATT-
40 GGGGAGAGGTGGGGCAGGGATCCTGGTAGAGACCAATGTTTCCACCCAGACCC-
CAAGACTGCTGGGAGAGATGGTGTGAGCAGTGACTCCAGGAATATCCAGTGGTGTG-
GTGGCCCATCCAGGCCCGGCTGGGCAGGTGGCTGGCTTGGTGGGGATGTGAT-
GATGGTGGTAGGCATGGGAGGCACCTTGGACGGGATCTGATTTGGCAAAAGGAAGTG-
GTTTCCTGTCCCCAGTGATTTCCAGCCCTTCCAGACCTCCCAAGGCTAAGGCAGAT-
45 TACTAAATTTAAGGCTGGGGCCCTCCTTCTTCCCTGGACTTCCAGGAGAACAGAGAAC-
CGGTGGCAAGGACCACCACCAGCAGGGTGAGGGGTGCAGATAAAGGCAGCAAAAAA-
CAGAGGGAGAGGTCTGGAGGGAAGGCAGGAATGCTTGTTCGTGAGCCTCAGAAAC-
CTCCTTCTATCCTGCTAGACTTTACTCCTTTGAGGCTTACCCTGGGGAACAGCTGGG-
GAGAGACAGGATCTTCAGACATCAGGAGCTCCACCTCCTCATCCACATGCAATC-
50 CGCTGCCTGTCTATCCTCCACCCCTTCTAAGGGGACCTCTCAGCACCTCC-
CAAAGTGCTCCAGAATCCAAGTTCTGTGTACCTCCAAGAACAGATGGAACCTTCCA-
ATCAGAGCCTCCACTGATGAAATGGAATATTTCCAGTGCTCTCTAACTGCCATAAGGA-
GAAGCCCACCTCTCTCTAACACCTTGGTTGCTTTTTGGGTCCACCTCCATATT-
TAAAAAATCTCCTCTCTCAGGGCCGGGAGCAGTGGGTACACCTATAATCCCAGCAGT-
55 TTGGGAGGCCGAGGTGGGTGGATGACCTGAGCTCAGGAGTTCAAGACAAGCCTGGT-
CAACATGACGAGACCCTGTCTCTACTAAAAACAAAAAATTAGCTGGGCGTGGTGGT-
GCATGCCCGTAATCCAGCTACTTGGGAGGCTGAGGCAGGAGAATCACTGAATCCG-
GGAGGTGGAGGCTGCAGTGAGCCAAGATCGCGCCACTGCACTCCAGCTGGG-
CGACGCAGCTGAAGCTGTGTCTCAAAAAACAAAAACACACACACACACACACA-

P 687 DK01

49

5 GAAAAAAAAAACCAAAATAAAAAATCTCCCTTCTCAGGAATGTAACGGAATCTTCCTT-
GCCTTCTCCCCTAACCTAATAGAGAATTTTCTCAGTTACACTGTAATTTTATTAATG-
GATTTTCTCCTCATTCTGCCAATGCAGTGTAAAGCTTCTCTCCATCTGTTATAT-
TATATATAAATATATATTATATATTTATATATTATATATTATATAACATATAATTT-
TATTGTCACCCAGGCTGGAGTGCAGTGGCACCATCAGGGCTCACTGCAGGATCAATC-
TCCAGGCTTAAGCGATTCTCCTGTGTACGCTCCTGATGAGCTGGGATTACAGG-
CAGCCGCCACCAACCCGGCTAATTTTTTTTTTTGATTTTTAGTAGAGATGGAGTTT-
10 CACCATGTTGGCCAGGCTGGTCTAGAATCCTGACCTCAGGAGATCCGCCCGCCTT-
GGCCTCCCAAAGTGCTGGGATTACAGGTGTGAGCCACCTGGCCGGGCCCTCCACT-
TCCTTCTTGTACATTGCTGAATCCCTGTGTACGCCCTAGAGGTCCAGTCTTTTGCCTC-
TCCAGCCTTAATCTACAATTCTGTAACCCACCCACCATCATTAAATGAGATTCTTCT-
TTGTGCTTCCCTTGGCTAAATGGATTATCTTTAACCTCTCCACCAATACAACCAGG-
GATGATAATAAAAAACATTGGATTGAGCAGAAACCAATCAAATAACTAGTAAGGCAGTAC-
15 TGGCGAGCACCTACATCCTGACAGCTTTATAAAGGGCGCTTCCAGCCAGGTGCGGT-
GGCACATGCCTGTAATCCAGGACTTTGGGAGGCTGAGGCGGGCAGGTACCTGAG-
GTCAGGAGTTCAAGACCAGCCTGGCCAACGTGATGAAACCTGTCTACACAAAATA-
CAAAAAAAAAAAAAAATTAGCCGTGCGTGGTGGCATGCGCCTGTCTACACAAAATA-
TCTGGAGGCCAAGGAGGGAGGATCACTTGAGCCCGGGAGGCAGAGGTTGCAGTGAG-
20 CCCACATCTTATCACTGCACTCCAGTCTGGGTGACAAAGCAAGACTCCATCTCAA-
ATAAATAAATACAAATTGGCCGGGTGCGGTGGCTCATGCCTGTAATCCCAGCACTTTG-
GGAGACCAAGGCAGGTGGATCATTGAGGTGAGTAGATCAAAACCAGCCTGGCCAA-
CATGGTGAAACCCCGTCTCTACTAAAAATACAAAAAGTAGCCGGCGTGGTGGTGGT-
GGGCGCCTGTAATCCAGGCAGGAGAATGGTTGAGCCCGGGTGGGGGGGCGC-
CGAGGTTGCAGTGAGCACAGATGGCGCCATTGCACTCCAGCCTGGGCGACAGAG-
25 CGAGACTCCGTTTCAGAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA-
ATTTAAAAATAAATAAAGGGCTTTTCTCAGCTTCTCAGCTTCTCAGCTTCTCAGCT-
TACCCCGACATTACTATTAATATAACGGACTTTTCTCCTCCCATGAGCAATA-
ATGAGCTTTTTCAGACCTCCCTCTCCCAATATAACGGTTTGTCTCTGTTGCCTCTTCTT-
TTCTGTGGGATCCCCCTTTTCCCAACCCCAACTGTCCGGAGGTCCCCATGACTTC-
30 TCCCTGGGCTCACCCCGAAGTAGTTCCGCGGCACGTAGCCCTCCTGGCCGTGCAG-
CGCGGCCACCACAGTGGTCTCCTCCGGCCCGTCCCTCCGCAGCACGGTGAC-
CGACTCGCCCTCGCGGAAGGACAGCTCGTCCCCGAACCTCGGCGCTGTAGTCCAGA-
GAGCGTACACTGCCCGCTGTTTCATCAGCCCCATACTCTGCTCGACGTCTGAAACAT-
GCCACGGAGGGGAAGGTGAGAGCCTGCCAGGGGGTCCAGGAACAGGGGGC-
35 CACGTGGGGTCCAGGACAGACCTTGGAAATTTGGCGCCTGTCCAGCAACCACTGAA-
ATGTTGTGTGTGCCATGGCTGTGGATGGGAACCGGAGCTGGAGTCAGATGCCGG-
GACTGGCCGTCTTTGAGCGTTTCAGGAAACTGGGGGAGGCATGCCAGTGGGCCACC-
CACTCCCGAGGCAGGGTCAGAGGCTCCATTCTTTCTTTCTTTTTTTTTTTTGA-
GACAGAGTCTCGCTCTGTGCGCCAGGCTGGAGTGCAGTGGCACGATCTCGGCTCAC-
40 TGCAACCTCCGCCTCCCGGGTTCACACATTCTCCTGCCTCAGCCTCCCGAGTAGCT-
GGGACTACAGGCGCCCGCCACCACGCTGGCTAATTTTGGTATTTTTAGTAGAGT-
CAGGGTTTACCCTGTTAGCCAGGATGGTCTCGATCTCCTGACCTTGTGATCCGCC-
CACATTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCAGCGCCCGGCTTT-
TTTTTTTTTTTTTTTTTGGAGATGGAATTTGCTCTTGTGCGCCAGGCAGGAGTGCA-
45 ATGGTGCGGTCTCACTGCAACCTCCGCCTCCGGAGTTCGAGCCATTCTCCTGCCT-
CAGCCTTCCAAGTAGCTGGGATTACAGGTGTGCGCCACCATGCCTGGCCAATTTTTG-
TATCTTTAGTAGAGACGGGGTTTACCATTGTTGGTCAGGCTGGTATCAAACCTCTGAC-
CTCAAGTGATCCACCCGCTCGGCTCCCAAAGTGCTGGGATTACAGGCGTGAGC-
CACCTGGCCCGGCCCTCATTTCTTCTTGTACATTGCTGAATGCCCGTGTCAACCCTA-
50 GAGGTCCAGTCTTTTGCCTACCCTGGCGCTTAGCTTAAGTGGTACAGTCTCTAAG-
GAAGATTGACACCTTCTTGAATGATAGGGTCTTTAAGTTGGCTCATCTGCCTCTTTC-
TTTTCTTTCTTTCTTTCTTTTGGAGACGGAGTCTTGCTCTGTGCGCCAGGCTG-
GAGTGCAGTGGCGGATTTCCGGTCACTGCAACCTCCGCCTCCTGGGTTCCAGCAAT-
TCTCCTGCCTCAGCCTCCAAAGTAGCTGGGACTACAGGCCACGCGCTACACCCGG-
55 CTAATTTGTTTTATATTTTAAAGAGACGGGGTTTACCCTGTTGCCAGGCTGGTTT-
GGAAATCCTGAGCTCATGCAATCCGCCCGCTCGAGCCTCCCAAAGTGCTAGGATTA-
CAGGCATGAGCCACCGCGCCTGGCTTTCTTTTCTTTCTTTTCTTTTCTTTTCTTTT-
CAAGGTCTCACTCTGCCACCCAGGCTGCGGGAGTGCAGTGGTGAATCAAGCTTACT-

P 887 DK01

50

GCAGCCTCGAACTTCCAGATTCAAGCAATCCTCCTGCCTCAGCCTCCTCCTGATTCTT-
TATGTTATTATTAATATTTTGTAGGCCGGGCACAGTGGCTCAGACCTATAATCACAG-
CACTTTGGGAGGCCAAGGCAGGCGGATCCTCTGAGGTGAGGGGTTTGAGACCAGCC-
TGGCCAACATGGCAAAACCCCGTCTCTACTAAAAATACAAAAAAGT-
5 TAGCGGGCCGTGGGGCCCTTGCTGTAAATCCAGTTACTCGGGAGCCTGAGGCAG-
GAGAATCGCTTTCACCGAGGAGGCAGAGGTTGTAGTGGGCTATGGTGCCATTGCAC-
TCCAGCCTGGGTGACAGAGCAAGACTCTGTCTCAAAAAATAAATAAATAAATAA-
ATAAATATTTTCGTAGAGGTCAGGTGTGGTGGCTCAGACCTGAATCTTAGCACTTTGG-
10 GAGGCCAAGGTGGGCAGATTGCCTGAGCTCAAGAGTTCGGGACCAGCCTGGGCAA-
CACTGCAAAACCCCTTCTGTACTAAAAATACAAAAAATGAGTCGGGCATGGTGGT-
GAGCACCTGTAGTCCCAGCTACTCAAGAGGCTGAGGCAGAGAATTGCTTGAATCCAG-
GAGGTGGAGGTTGCAGTGAGCCGAGATTGAGCCACTGCACTCCAGCCTGGGTGA-
CAGTGAGACTCTGTCTCAAAAAATAAATAAATAAATATTTGTAGAGACAGGGGGTCTC-
TACAATGTCTTGTAGCCTGACCAGGCTCACCTTTCAAATATATAACCTCTGTCTCACC-
15 CATAAGTCCTAGGACCTGCCTCACTCCAACCTCTCCGTGAAGTTCCTTGCCACACCCGA-
GATACAACTGGCTCCTCCAGGTGTGAAATGACCCTGTGCACAATCCCCGTGGCAGAG-
CCTACTTGGCCCTGCCCGTCCGGGAACAGGTGATGTAGCCTGCCCGCTGGAGAGA-
TAGGGTACAGCCTTGTGTCTTCTACAAGCCCTTTCTGGCAGCTGTAGCCTGCTCAC-
CTGCCAGTGGTGTGGCAATGCCTCTCCACAAGTGGCAGAGCCCACCTGCCAGAG-
20 CCCTATGCCAGGTAGATGGCAGGGTTGAAACGTTGAGCTCCTCACCTTGAAGATGT-
GAAAGGTGAGCAGACCAATCTTACAGCCACTCTCCTCCCCAAAGGTGTCCAGCTCG-
CATAGCAGAGCCTCCATGTCCCTTTTCCCTTAGGAGGGCATAGTCCCCCAACCC-
CGCAAGCGGTCCATCCCTCATCCTCCTCGGCAATCCTGCCAAGTGTTGGTA-
CAGCCCCATACCCTTCTCTCCCTAGTAGGGGGTAGTTGCTCCCCTCCCCGCTCCTG-
25 CGCAACCCGAGGTACCCAGGCGCCAGCAGCCCTGCCTCGCACCTGCCAGGTAGGT-
GGCGCAGTCAGCATAACCCCTCGCGGTAAAGGGTCCGACTTCTCGAAGGCGGTGGCGC-
CGTCGCTGAGCGTGGTGGCGAAGATTGCAGCGCCGTGCTGCACCAGCGCCATGCA-
GATGACTGTGTGTTGCACGACGCCGCGCAGTGCAAGGGTGTCTAGGCGTGGGG-
GTGGGGGGTTCGGGGGAACGATGCGTGAGAGGCTGCGCGTCCGCCACCGGGGAC-
30 CCAGCCCACCGCGGGTTCGGGGCTCACAGCCGTGGCTGTGGGGGAGTTGA-
CATTGGCACCCGCGGTGATGAGGAAATCCACGATAGAGTAGTTGGCGCCGAGAT-
GGCGTTGTGCAAGGCAGTGATGCCCTCCTCGTTGGGCTGGCTCGGGTGGTTCATCT-
GAGTGCACCGGGGGAGGGGGAAGACTCAGTCCCGCGGCTGGCATCTGCGATGCCC-
CGCGCGCGGGGACCTCCCGCTCAGCGCCGCTCCAGGAGGAGCACCAGAGGGTTGAGG-
35 CACCTCCAGCTCCCCGCTCAGCGCCGCTCACCTCCTTACCGCCTGCTGCAC-
CGCGCGCGGGGACCTTGGCGGGGAGCCCGCCTTCCGCAGCACAGAGCGCATC-
TCTTGGGGGACAGGGCGCAGAGGTGAGCGACTTGGAGGGATTGTTAGTATATCCAT-
GATCTAGAGTAGGAAACAGAGGTCCAGGGACTTGTGGCACCCATCTAGACAGGGGTA-
GAACTGGGATTCCCTCGGGATGGGGTAGGGGGTGCCTTCGATCTCCTCTAGAGCC-
40 TCCAGTTCCTTCCATAGACAGGGAATCCTGTGATTGAGAATCTTGGGCCCTGAAAC-
TTGGGAGAAAGCTGGGGGGCCATGGGATTGGTGCCAAAGTAATTCTATCAGTT-
CAAAACAATGATTGTGGAAGCCAGTTATGCAATTCACACACAGTCTCACATTTCTTTT-
GTTAATAATGAATGCAATGAGACACACATGACAAAATGTTACCAGGAGTGTTCATTC-
CGGATGTTTGAATTTGAGCATTTTATTATTCCTTGTATTTCTTTTCTTTTCTTTT-
45 TTTTTTTTTTTTGGAGATGGAGTCTCGCTCTGTCAACCAGGCTGGAGTGCAGTG-
CAGTGGTGTGATCTCAGCTCACTGCACCTCCATCCCCAGGTTCAAGCAATTCTCCT-
GCCTCAGCCTCCTGAGTAGCTAGGATTACAGGCATGCGCCACTATGCCTGGCTAATTT-
TCATATTTTAGTAGAGACAGGGTTTTGTGATGTTGTCCAGGCTGGTCTCGAACTCCT-
GACCTCAGGTGATCCACCCACCTCAGCCTCCCAAAGTGCTAGGATTACAGGTGTGAG-
50 CCAGTGTGCCAGCCTCATGGGCTTTCTATTTTAAATTTCTCCTGTAAGATTCAAT-
TATTCTGGGCTGGGCGAGGTGGTCTGTGTAATCCTAGCACTTTGGGAGGCT-
GAGGTGGGAGGATCACTTGAGCCAGGAGTTGAGAGAACAGCTTGGGCAATATAGTGA-
GACCCAGTCTTACAAAAATAAAAAATTAGCCTGACATGGTGGCGCACACC-
CGTGGTCCAGCTACTTGGGAGGCTGAGGCAGGAGGATTACTTGAATGGAAGAGAAG-
55 GAGGCTTCAGTGAGCCATGATGCCACTGCACTCTAGCCTGGGCAACAGAGTGA-
GACCCAGTCTCAAAAGAAAAAATGCAATTTATTTCAAAGTGTGTGAGTGCATAG-
CATTTGTGATTCTGGTCTTTGCTGTTCCAGAGTTTCAGTGATTTAAGATTCTGGAATT-
CAGAGATCCCAACAGCCACTGAATTCAAAATCCAGATGCTCAGTTATTCAAGTTTC-

P 687 DK01

51

CAATATGTTGTGATTGCAGAAATGCTAGGCTGTGCTATTTCAAATTGCTGAGGGGC-
CAGGACTTTGGAATCCAAAGATTCTATGATGGAGAACTTTAATATTTTCTGTTAGAATT-
TCTTTTTTTGTTGGTTTTTTTGAGACAGAGTCTCGCTCTGTCGCCCAGGCTGGAGTG-
5 CAGTGGTGCGATCTCAGCTCACTGCAAGCTCCGCTCCCGGGTTCCAGGCCATTCTCC-
TGCTCAGCCTGCCAAGTAGCTGGGACTACGGGCGCCCGCCACCAGCCTGGCTATT-
TTGTATTTTTAGTAAAGATGGGGTTTCACCGTGTAGCCAGGAAGGTCTTGTCTCT-
GACCTCGTGATCCGCCCACCTCGGCCTCCCAAAGTGCTGGGATTACAGGTGTGAGC-
CATCATGCCTGACCTAGAATTTCAATTTAAAAGACTAGAAGGAAATGGCTGGGTGCG-
10 GTGGCTCATGTGTGTAATCTCAGCACTTTGGGAGGCTGAGGAGAGTGGATCACCT-
GAGGTCAGGCAGGAGTTCAAGACCAGCCTGGCCAACGTGGTGAAACCTGTCTCTAC-
TAAAAATACAAAATTAGGTGGCCGTGGTGGTGACGCTGTAAATCCAGCTACTCAG-
GAGGCCGTGGCATGAGAATCACTTGAACCCAGGAGGCACAGTTATAGTGAGCTGA-
GATGGCACCATCGCACTCCAGCCTGGGTGACAGAGTGAGACTCCATCTCAAAAAAG-
15 GAAAAAAGAAAGACTAGAAGGAAATATTCAAATGTTAATGATGGTCCCTGT-
GAGTGGTGTGATTTTGTCTCTTTCTTCTATTTTTATTTATTTTCCCCAAGCTCTCTATG-
GTGTTGGTGTATTTCTCTATAGTGGAATGTGTAATTTAAAGTATAAATCTCAGCTGGG-
CACAGTGGCTCATGCCTGGTTTGAGACCAGCCTGGACAACATAATGAGAACTGTCTC-
TACTGAAATGTTAAATATTATCTGGGAGTGGTGGTGCATGCCTGTAGTCCCAGC-
20 CATAGGGGAGGCTGAGGCATGAGGATCAATTGAGCCCAGTAGGTGGAGGCTGCAGT-
GAGCCATGATCTTGCCACTGCACTCCAGCCTGGGCAACAGAGTGAGACTCTGTC-
TCGATAATAAACCCTCTATTACAACATACAGTGCATGAATTTGTGATTTTATAATT-
CAAAATATGAGCATCTTTAATTGTGATTTGGTGACTTCAAGAATCAGTAATAAT-
CAGTCTATGATACTAACTTTATAATTATTTTTTAAAGAGAAGAGTTTCCCTTTATTT-
25 TATTTATTTGAGACAGAGTTTCTCTCTGTTGCCAGGCTGGAGTGCAGTGGCGCA-
ATCTCGGCTCACTGCAGCCTCTGTCTCTAGGTTCAAGCAATTCTCCTGCCTGAGCC-
TCCCGAGTAGCTGGGATTACAGGCATGCACCACCAGGCCAGCTAATTTTTGTATTTT-
TAGCAGAGACGGGGTTTACCATGTTGGCGAGGCTAGTCTTGAACCTCTGACCT-
CAAGTGATCCACCCGCCTCGGCCTCCCAAGGTGCTGGGATTACAGGCATGAGCCAC-
CGTGCCCAGCCTAACTTTATAATTCTAAGATCGTGTTCAAACCTTTAAATGCTCTAGGG-
30 CTCTAAAATGTTACTATCCTAAGACGTTGACACTAGCGTTTGATTCTTACATTCCTATGAT-
TTTTAAGTTTCTCTGTGGCCAGGACTCTGTGATTCTACAATGGGATGCTCAGCCATTT-
CAACATGTTGTTATTCATCCCCTCTTGATTTCAAATCCTGAGCCTCAAGGTTCCCTTGC-
CTTTACTTTCAGGAGGGCCTAGGAATAGGCATTTTGGGGGGGTCCACCTGACCCCTG-
CTTCTCTGAGAAGTGATCTCTTCCGCTGTCTACGCACACGGAGTGTTCAGGACTGT-
35 TCCATGTGGCTACAACCCCTCTTCCAGTCAAGATGCAGGGACCAAGATCAGCAGGA-
GACCATCCCCTGGTCCAATGGTGACAACAGTAAGAGCAGTTAACAGTTATGTGCCAGG-
TATTATGCTAAGCACTACATTAATGTATTTAATCTTGGCGGGGTGTGGTGGCTCACACC-
TGTAATCCCAGCACTTTGGGAGGCCAGGGCGGGCAGATCACTTGAGGTGAGGAGTT-
CAAGACCAGCCTAGCCAACACAGTGAACCCCATCTCTACTAAAAATACAAAAATTAG-
40 CCAAGCGTGGTGGCATATGCCTGTAATCCCAGCCACTTGGGAGACTGACGCAGGAGA-
ATCACTTTAACCAGGAGGTGGAGTCCAGCACCCAGCCGAGACTCACTTGTTTTATT-
TATTTATTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT-
CAGGCTGGAGTGCAAGTGGCGGATCTCAGCTCACCACAAGCTCCGCTCCCGGGCT-
CACGCCATTCTCTCTCAGCCTCCAGAGTAGCTGGGACTACAGGCGCCCGCCACCAC-
45 CCCAGCTAATTTTTGTATTTTAGTAGAGACGGGGTTTACCCTGTAGCCAGGATG-
GTCTTATCTCCTGACTTCGTGATCCGCCCGCCTCGGCCTCCCAAAATGCTGGGATTA-
CAGGCATGAACCACCACGCCCGGCCTATTTATTTATTTATTTATTTATTTATTTATTT-
TCTGTGCCCCAGGCTGGAGTGCAGTGGTGCAGTCTTGGCTCACTGCAACCTCCGCCT-
50 TCCGGGTTTAAAGCGATTCTCTTGCCTCAGCCTCCTGAGTAGCTGGGATTGGAATGA-
GACCACCACTTCTCCTGTTGTCTTCCAGCTTCTCCCCACCTCCCTTTTCCCTAGT-
TTATAAGACAGGAAAAAAGGGAGAAAAACGCTGGAAAAAACAGAAATACGA-
TAAATAGCTAGATGACCTTGGCGCCACCATCTGCTCCTGGTGGTTAAATAATAATA-
ATAATATTAATCCCTGACCAAACTACTGGTGTATCTGTAAATCCAGACATTGTAT-
55 GAGAAAGCACTGTAAACGTTTTGTTCTGTTAGCTGATGTCTGTAGCCCCAGT-
CACGTTCTCAGCCTTACTTGATCTATCGTGGCCCTTTCAGTGGACCCCTTAGCGTT-
GTAAGCCCTTAAAGTGCTAGGAATTTCTTTTCGGGGAGCTCGGCTCTTAAGACGCT-
GATGCTCCCGGCCGAATAAAACCTCTTCTTCTTAAATCCGGTGTCTGAGGAGTTTT-
GTCTGTGGCTCGTCTGCTACAGAATTACAGGCACGGCCACCCTCCGGGCTAATT-

P 687 DK01

53

TTATCTTCAGTCACACAGCAAGTTAGCAGTGAAGAGAGACTCCAGCCCATCTGCT-
TAACTCACTGATCTCACACCTCAAAATATTAATAAATTATTATACTAATATGGTAGC-
TATTTATTTGAGACTGGGTCTCACTCTGTACCCAGGCTGGAGTGCAGTGGCGCTAT-
5 CACAGCTCACTGCAGCCTGGATCTCCAGGCTTAAATGATCCTCCACCTCAGCATCC-
TGAGTAGCTGGGACTACAGGCGCCCACTACCATGCCCGGCAGATTTTTGTACTTT-
TATTTTTAGTAAAGTCTATTTTAGTTTCACTATGTTGCCAGGCTGGTCTTGAACCTCA-
GAGCTCAAGCAATCCTGTCTGCATTAGCCCACCAAAGTCTAGGATTACAAGGGT-
GAGCCACGGTGCCTGGCTAATATGGTAGCTATTGATAGCTTACTATGTATCAGATCC-
10 TATTTATTTATTTTGGAGACAGAGTCTACCCCTGTACCTGTGCTGGAGTGCAGT-
GGCATGATCTTGGCTCACTGCCACCTCCGCTCCTTGGCTCAAGCTGAGTAGCTAG-
GACTACAGTGGTGAGCCACCATGCCAGCTAATTTTTTTTTTTTTTTTTTTTGATA-
GAGATGGGATTTTCATCATGTTGTCCAGGCTGGTCTTGAACCTCTGACCTCAAGTGATC-
TGCCACCTCGGCTCCCAAAGTCTGGGATTACAGGTGTGAGCAACTGCACCTGGC-
15 CCATCAGGTGCTGTTTTAAAGGCTTATATGAATTTAATAACATATGTCAATAG-
GATCGATTCTATCATTATTTGCCTTTTTTTTTTTTTTTTTTTTGGGAGAGTCTCCC-
CGTCAACCAGGATGGACTGCAGTGGCGCAATCTCGGCTCACTGCAACCTCCACCTCC-
CGGTCCTCAAGTGATTCTCCTGCCTCAGCCTCCCAAAGTAGCTGGGACTACAGGCGCC-
CGCCACCATGCCTGGCTAATTTTTGTATTTTAGTAGAGATGGGGTTTCATATTGGC-
20 CAGGCTGGTCTCGAAGTCTGACTTTGTATCCGCCCGCTCGGCTCCCAAAGTGC-
TGGGATTACAGGCATGAGCCACCGTGCCCGGCCATTATTTCCCTTTTACACTCAA-
GAAAATTGAGGCCAGTGAGGTTAAGTGACTTGCCCAAAGGTACACAGCGTGGAAC-
CAGGCAGTCTGGCTCAGGGTCCACACTTAACCTTTGAGCTATCCCTGGCTCCTACC-
CAAATTCCCAAAGTCACTGCTGCTGAGGAGAGTGGTGTAAAGAGT-
25 CATTGGCTGTGATCTCCCCACCTCCAGGCTGGTCTGGTCCCCCTGCCATTTGTCC-
TCCCTTCAACCAGTCTCTAGGGCCCTCATTGCTGACTCACTTCGTTACAGGGGC-
CATGTCTGTTGGGATGCTGGGGCTGGGGTAGGGTTTGGGGTGGGTCTGGGG-
CTGTGGGGGAGCTGGGGCTGTGGTTGTGATTGTGGCTGGGGCTGTGGTTGTGGT-
GGGGCTGCAGCTTAGGCGGGGGTGTGGGTGAAGAGGGGGGAGCCAGGGAGCAT-
30 GCGCGGCTGGCCCCGTGCTCCAGAAGGCGTCTGCAGCTTGAAGATCATGCT-
GAGGGGATGGGACGCTGGCGCGGGCCCGGGGCTGGGGGCTGGAGGGGGG-
CATGGGGATGCGGCTGACGGGCTGCCAGCTGCGAGGCAAGTGCCCGACGGCC-
CGCGGAGCCAGCGAGCGCGGTAGCTGCCCGCTCTGAACGCCGGTCTGCTGGC-
CAGAGGAGAGACCTTGTAAATTGCGCGGAGGGTGGCGCTAGTGAGGTTGCTCTGG-
35 GAAGAGGGAAGGGAGAAGGGTGGGTGAGAGAGGGAAGGTGGAGGGGAGG-
TAAAGACAAAAGACGAGAAGGGAGAGGTTGAGGGAAGCCCTGGGAGTGAGGGA-
GAAGAAAGGGTGAGGAAGGAGCAGAAACCCAGCACAGTGAAGGGAGAGCGTG-
GAACGGGCGCCGAGACCCAGATCGCAGCCCCGAGGGGGAGACTGGCCTTGACCC-
CGTCCCCCACCCACTCCTCGACCTTCCCCAGCCTCTCCTCCCCAGGCGTCCGCTC-
40 CTCACCTTGCCGGTGCCCCCAGTCCATCCAGGCTGCTCTCCCTCAAAGGCAACAGC-
TGCAGGCTCGGCGAGGCAGGCTTGCAGAGAGCTCCAGGCTGCGGGGCGGGAAT-
CATTAGGGTCTGTGGGGCTGCCTCTCCTCCGGTCTCCTTCCCCGGGCTCCAC-
CACTCACGTTTCATAGCTCGCTGTCTGCGAAGGCTTCTTCTCGTACGCCACGTCCAGGT-
CAGACTCGTTCCAGGCTTTCGGAGGCCGCGCGCAGCGTCAGGTGCTGTGGGA-
45 GAAGTTTCCAGGGAGGATGAGACGGGAGGGGTGGCGAGCCCCGATCCTGCCCGCT-
TTGACCCCGCGAGTCAAAGGCCCGCGAGGGGCCCTGGGTTACCTTGCGCGCG-
CAGAGGCGGGGGAATGCGCTGCCGCGGAGCCTAGCAGGGAGCTCCCGAAGGCG-
GACGCTGGCGCGTCTAGGCTGTGGCAGGGGGCGCGGTGACGGCCACGCTCGG-
GGAAGAAGGCTGGGGCCCTCCGCCAGGGGGCTGCCGCGGGGGAGCCTGGC-
50 CGGGCCGCGGGGAGGGCGCACGGCCGAGGGAGCTGCCTGCGCCATCGAAGGCG-
CGGGCCGCGGGGAGGTCGCGCGGTCCAGGCTGCCGTAGGCGTCCGGCTGCAGG-
TAGAGCGGGGTGCGCGGCGACGACGGCCGTCCTTGGGGACAGCGGGCTGTAGG-
GGTGTAGGGTTGGGGCACTCTCTGATCGTCCGAACGGGGTGTCTGCGCGTGGT-
GGCGGCTTCCGGGGGGACCTCGGCTGCCGAAGGGCTCAGGGATCGAGCTGGAG-
55 CTGTACCGGGGCGGCTGTGGGGAGGCCAGGGCATTGAGGGATGGATCAAAGGAGA-
CATTAGTGAAGGGTTGGTGTGTGGCGGGGTGTCAAGAGAGATCACTGGAGGT-
CAACCCAGAGGAGGCTGACCGGCCATGGAATTCAGGCACAGAGAGCCAGGTGAG-
TAGTGGTGGGGAGACAGCCCTGAATCAGCACTGTGGCTAGCCCATCTATGT-

P 687 DK01

54

5 CACCTTTATGCCACTTAGGTAACACCTCTTTCCTTCTGAGGGTCCCTTTAGATGTC-
CACTTCCACTGGTCCCCTCTTTTCTATTTCTTTCTTTCTTTCTCTCTCTTTCTTT-
TCTTTCTTTCTTTCT-
TCCCTCCCTGCTTGGCTTGGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT-
TCTTTCTTTCTTTCTTTCTATCTCGGCTCATTGCAGCCTCAACCTCCCTGGCTTAGT-
GTGATCCTCCCACTTCAGCCTCCCAAGTAGCTGGGATTACAGGTATGCACCACCA-
CACCTGGCTAACTTTTGTATTTTAGTAGAGACAGGGTTTACCAGTTAGCCAGGCT-
GGTCTTAAACTCCTGACCTCAAGTGATCCGCTGTCTCTGAAAGTGTTGAGATTA-
10 CAGGCGTGAACCACCGTGCCAGCCAGATTTTTAAAAATCATTGTAGAGGCTGGTC-
TCAAACCTTAGTCTCAAGCAATTCTCTACCTCGCCTTCCAAAGTGCTGGGATT-
CAGGTCTGAGCCATCGCGCTGGCCTGGTCCCCTTTTTCAAGTTCCTTGAAGAGC-
CCACAACCTGCATAACTATATGGGGCAATTTTGCTGAAATCCAGGCCTCTGGTCTG-
GACTGTGGCGAGAGGCTGGCTTTGGAGATCAAGGTGGGAACCAGGCTTACCCTA-
15 GAAGGGGGTCCGGCTGCGGGCAGGAGGCGGAGAGTCTGACCACAGCGAC-
TCCAGCTGCTTGGTCAGTTTATCCACCTTGGCCGCGCGCTGTCCAGCTCCATCTGC-
TTCAGATCCATGTGTTTATGGCCAGCGCTGGGAAGGTGGGAGTGGAGGTAAGGACC-
TGGCTCTCTGGCAGGGGGCGGCTCAGCACCCTCGCCGCTGCCGAGGTCCCCG-
CCTCGCCAGCCCCGCCCCCTACTCCAGCTTACACTGGAAGTTTATGTTCCAGAAAGTC-
20 CCGCGCGCTCTGGAATGCCTCGCTGTCCATGGTGCCGGCCGGAGCGGCGCTG-
CATGGTGGGGAGGGAGGGAGCTGGCTAAGACCCCGCCCCCTCTAGACCCCGCCT-
CAGGGAGTCAGACGCCGTGAGGAGCGGGACAACGCCTCAACTCAGTTCTTCCCT-
GGAAGCCCTTACCTTTTACCTTCCAGCTGGGAAATGCCAACTCCTCCAAAGC-
CAAGTCCATGCGCCACGGAGAAGTCCAAACCCAGTCTAAAACCTCCGGAATTCATT-
TCTCTTTCTTTTTCTTTCTTTTTTTTTTTTTTTTTTGTGTATGTGTGTGAGACA-
25 GAGTCTCGCTCTGTGCGCCAGGCGGGAGTGCAATGACCGCATCTTGGCTCACTG-
CAACCTCCGCTCCCGGGTTCAAGCAAATCTTCTGCCTAGCTGGGACTACAAGCGCG-
CGCCATTATGCCCCGCTAATTTTTGTAGTTCTGGGATTACAGGAGTGAGTCTCCGCG-
CCGGCCGTGTCCATCTCTTATCTCAGTCTTAAGACCTGAATCACTCTTGAACAAT-
TATCTATTGATCACCTACAATGTGCCGGTAAACATAGGATGGAATAACTATGAATTACT-
30 GAATGTTTACTAGGGACGACGCACTGTGCTAGATCCTGTTTTGTTGTTTTGA-
GATGGTGTCTCGCATTTTTCGCCAGGCTGGAGTGCACTGGCGCGATCTCGGCTCACT-
GCAAGCTCCGCTCCAGGGTTTATGCCAGTCTCCTGTCTCAGCCTCCCGAGTAGTG-
GGACTACAGGCGCCTGCCACCATGCCTGGCTAAATTTTTGTATTTTAGTAGAGACGG-
GGTTTACCCGTGTGAGCCAGGATGGTCTCGATCTCCTGACCGCGTGATCCATCTGCC-
35 TCGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCGCGCCCGCCTTGT-
TTGTTTTTAATAATAATTCTGCTGTCTGCTGTGTACTAGAACCCTATGCCTACTGTT-
GGGTATAATGTAGTAAATGTAGTAAAAACAATATCCGCCGGCGCGGTGGCTCACGC-
CTGTAATTCAGCACTTTGGGAGGCCAAGGAGGGCGGATCACGAGGTGAGGAGAG-
CGAGACCATCCTGGCTAACATGCTGAAACCCCGTCTCTACTAAAAATACCAAAAT-
40 TAGCCAGGCGTGGTGATGGACGCTGTAGTCCAGCTACTCGGGAGGCTGAGGCAG-
GAGAACGGCGTGAACCCGGGAGGTGGAGCTTGAAGTGAAGCGGAGATCGCGCCACTG-
CACTCCAGCCTGGGCGACAGTGCGAGACTCCGTCTTAAAAACAACAAATAATAA-
ATATGTTTAAAAACAACAACAATAAACCAGCCAGGCGCGGTGGTTCACTCCTGTAA-
45 CCGAGCACTTTGGGAGGCCGAGGTGGATGGATCGCTTGAAGCCAGGAGACCAGCT-
GGCCAATATGGTGAAACCCGCTCTCAAAAAAATACAAAAGTTAGCTGGGCATGGT-
GGCATGTGCTGTAAATCCAGCTACTCAGGAGGCTGAGGCACAAGGCTCACTTGAAC-
CTGGGAGGCCACAGGTTGCAAGTGAAGCATAGATTGTGTCACTGCACTGCAGCTTGGGT-
GACAGAGCGAGGCTCTATTTAAAAAATAAATTGAGGGGCCACTCCCTCTA-
50 GAGTGGTGAGAAATGCCGTGCACCGAAAGCTTCAATTTGATGGTCAAAACCACTTAG-
CAGGCAAGAAAGCATGGCTCAGAAACATATGTTCAAGGTCAACCTGCAAGAAGTCGG-
TAGTAATCGGTTTACACCCGCATCTAACTTATTCTGGGTCTCTCTACCAGATTAGAG-
GGGTCTAGAGGGAAGCGACTGCTCAGCTTCCCTTCCCTAGGGTCCCATTCACTG-
GAGGTCTGGCTCTCACTGACCCATTGTTAGCAAGAGGAACAGGGAGGTGGCCAGGG-
GTGGAGGGGCGAGCTGTGGTCACTGGCCAGTGGGAGGGAGCTAGGCCACTAGGAAC-
55 CGGTGAGGCCAGCACCATCCCTATCCCATGCTAGCCACCACACCCACAGCTCTGC-
CACCTCCCTGCTGCATCGACCACTTAGCTCTGGCAGTATAGGCAGCAGGGCAGGCTG-
GGGCATGCTGATACCCGCTCTGTCTGGGAAGTCGAAGGAACAGAACCTGTTCAAGG-
TGCGCGCTCATTGGATGAACAGGGAGTGTGTGACCTTGGGCGTTGAGTCTCTC-

P 687 DK01

55

CACTCCCTGGGCCTCAGTCTCCCCAACATCAAAGAAGAAGGCAAATCACCTTTTTTTTTT-
TTTTTTGAGATAGGGTCTCGCTCTGTAACCCAGGCTACAATTGTGACTACTACAGCC-
TCTTGACCTCCCAGCTCAAGTGGTCCCTCCACCTCAGCCTCCTGAGTAGCTGAGACTA-
TAGGTATAGCCTCGCACCACCACCCAGCTAATTTTTTTTTTTTTTTTTTTTTTTTTTTT-
5 TTTTTTTGAGACGGGAGTCTTGCTCTGTGCGCCAGGCTGGAGTTCACTGGCGGGGATC-
TCGGCTCACTGCAAGCTCCGCCTCCCGGTTACAGCCATTCTCCCGCTCAGCCTCC-
CAAGTAGCTGGGACTACAGGCGCCCGCCACTACGCCCCGGCTAATTTTTGTATTTAG-
TAGAGACGGGGTTTACCATTTTAGCCGGGATGGTCTCGATCTCCTGACCTCATGATC-
10 CGCCCGCTCGGCCTCCCAAAGTGTGGGATTACAGGCGTGAGCCACCGCGCCCGG-
CCACCCAGCTAATTTTTTAAAAACATTTGTACACTTTGGGAGGCTAAGCGGGGAG-
GATCACGAGGTGAGGAGCTCGAGACCATCCTGGCTAACACAGGTGAAACCTGTCTC-
TACTAAAAAATACAAAAAATTAGCTGGGCGTGGTGGCGGGCGCCTGTAGTCCCAGC-
TACTCGGGAGGCTGAGGCAGGAGATGGTGTGAACCAGGGAGGCGGAGCTTTCAGT-
15 GAGCCGAGATCGCGCCACTGCACTCCAGCCTCGGAGACAGAGCGAGACTCCGTCC-
CAAAAAAAAAAAAAAAAAAAAAATTTGTAGAGACAGATCAAGTCTCACTTTGTGCTCAGG-
CTGGTTTTGAACTCCTGGGCTCAAGCAATCCTCCCGCCTCAGCCTCCCAAAGTGT-
GAGATTACAGGCATGAGCCACCACACCTGGCCAAATCAGCTATTCTGAAAGGCCCTT-
TAATCTCTATGAGCCCCAGACTTTCAAAGTGAAGGACCTTAGGACTGTAACATAAG-
20 TCTACAGAGCCTAAACCCCTCAGCTAAAGAGCCTATTGTTGAAAGTTCTGAGTCCAA-
GATTCTATCTTTGGAACATTCTAGAATTCTCCAATTTGTCTAACCAGAACTCTGAGTCT-
TTCTGTACCACATTCTACCTAACCAGGGTTGCACTGCTCTGGAAGTCTAGATGGATG-
GTATAGTGCAGCTGGTAAAGCATGAGTAAGAAGTCAAGCTTCAAAAAATCAAATCT-
GAGGGCCGGGCATGGTAGCTTCTGCCTGTAATCCTTGCCTTTGGGAGGCGGAGGG-
25 GGGAGGATCACTTGAGGCCAGGAGTTCAAGACCAACATGGCCAACACAATGAGAGCC-
CATTTCTTAAAAAAATTAATAAAATCATCAAATCTGGCAGCACCACCGTCCAACCC-
TGACCACAGTACCTCAGTCTCGTAATCCGTAAATGGGGATGAAAGTTCACCTCATAG-
GACTACTGTAAGAATCCACCTGGTCAAGGTCAGGAAGAATTCAAGCTCTGAGA-
ATTGAGGCCTCAGGAAGAAGAGACTACAGGAATAAAAACTCGGGCATTAGAATTTCA-
GAGATACACAAACAATACTTTGTTAACTGTTAAATAGATAAATGAGCAAGTCTGTG-
30 CAGCCCTAATGCCAGCTGTAAGTACTTTTTTTTTTCTTTGGTAGAGATTTAGTCTC-
TCTCGCGCCTGTGGTTAGGCTGGTCTCGAACTCCTAGCCTCATGGGATCCTCCCCGG-
CTCGATCTCCCAAAGTATTGGGATTACAGGCGTGAGCACGGCGCCATGATCCCCAA-
ATTTCAAAGATTCTCAGATTCCATACTGACATTCTCTGGCTCTCAGGAAATGCCAACCC-
TGGGTGTGGGCTGTGCGGGGACAGGCGGTGGGGACGTCGGAGCCACCAGGGGGG-
35 CGGTACGCCCCGACCCCGCCGCGGAGGCGGACTGCGCCTGAGCTCAGGCCCGG-
GGAATGCGCAGCGGGCCCCGGGCAGGTGCTGTACATCCCGGGGCAAGGGAGCTGGG-
CCGGGCGGGGTACAAGGGCGGGGCGCGGGGGTGGCGCGGGGCCGTGTGTCTGTTCC-
CAGGCCTCTGCCCTGACCTCTGCCTCCGAGTCTCTCCCATGTGCTCCCTCTAGC-
40 TCTAGCTCCGAGCTCTCCCGCGGGCTCTGGCCAGCCGAGGTACTCTCCCTGGG-
CTCCTCTCTCCGCTCCACCCCTGGCTCTCCTTCCCTGGCCTCCTCTGCACCCACAG-
CAGGTTCTTTAGGGCTAAGGATCCTGTGGACTTCTGGAGGAGTCACTTTCAGTAG-
GAACCGGGTCAGAGAGCCAGACTGAGCTGGGAACACCCAGGCTGGACTCCTACAGC-
CCTGTGCGGTCACACTGAACTCTGGAGAGGCTCCACTGTCTCTGGGACTCGGTTTCC-
45 TCTTTGTGGACGTCTATGGAATGGGCTAGGGCCTTTCTTGCTCTAAGCCTCTACTG-
GGCTTGTTATTTAGCTTCTCTGTGCCTGTTTCTCATGTGGACCATGGGAAGAATTA-
ATACCTTCGCCTCAAAGGGGTATGAGGATTGAGTGACATAATTTATAAGCCGTGATTA-
GAACAATGCAGTGCGCGAAATAAAGTTCACACATACAGGATTCAATATTACAGAT-
GTCCTTGGCTGTTCAATTATAATAACACAGGGTCTGGCAACAGAGTGAGGGGTCCAGAC-
50 TCAATGTAATTTTTTTTTCCCTAAAAGGGCCCTTTCAACTCTTCTGAGATCATACAAG-
CCCTGAGTTTTGACACCCAGGGTCTCAACTTCTGAGCCCTTGCCCTCTCAGAGTCC-
TAAATTTCCCTGTACATTCTGAGTCTGGCCAGTGATCACCTCAGTCACTTAGG-
GACGGGAGGGCTGGGAGAGCCCTGGAAGATTCCAGACAGAAGCTGGCAAAAGCC-
CAGGGTGTGGGCAATATCCACTCTCCAGCCTCCGTTTCTCACTCTAATGAG-
55 GAGTCTTCCCTGGGGTCAGCAAACCTTATTCAAAGGGAGACCTCTCAGTCAACCAA-
GATTCTCTAGACAATGCGAGCTTCTTACCTACCTACCTACCAGCTCTGAGCTTGG-
TACACCCAGAGCCCTGTTTTGGCAACAGGTTATTATTTTAAATTCATTTCAAGGT-
TATCATCAAATGCCCTTCAAGCCCAGACATTGGGAAACACTCCTCTCTCAGATGC-
TCGCCTCCCCATTCTGTTTTAATCCCCCTCTTAGGACGCATGGGGGTTGAGA-

P 687 DK01

56

GAACGGGGAGATAGACAGAGGGAGGTGCCTGGTCCTGCCCTCCCCCGCCTCAAG-
GACAGACAGACACCTCCAGAATTAGCCTCTGTCCCTCCTTATCTCCCACAATACCC-
CAGGTCAGACAGATGGGCGTGGAGGTGACATTTCTCACCTCAGGGTCAGGGCAAG-
5 GAGCCCTGAGGCAGAAAGTTAGTCAGAAAATCTGGCGGGGGCGGATGGAATCC-
CGTCCCCCAGAGAGCTGCAGAAGAAGGAGGAGGCAGAATCCTGACCTACAAACTC-
TACTGCCTGTGTGAGCTCCAAGCCTCAGTTTACCCCTTCTCTCCGTGTAATGGTTAA-
ATGCCCGGCTATGCAAACCTCCAGAAATCCAATAGCGCTTTCGGGAATTCTGCCCT-
GGGTTCTAGAACTACCTCTGCAAACCCAGCTGTTTCCACCCCATAGGGCAATAGGG-
10 GAGCCACCTCCGCCAGGGGGTGCCTAGGGCGGATGTCCCTTCTCTGGTTAGG-
CAGGTCTGACGCCAGGTTAATGACATGTTGGGTTGCTCAGCGGCACAGAGGAG-
GTTGGAGATCTGCCTCGGTGTTTTCTCTCTACCCCGCCCCATCCCCGAGC-
CGAAAAGTCGGGGGAGAGCCGGGACACAGCCTCCGGAGGGACCCCGGGTACCT-
GTCTGTCTCCACTTCAGGAACCCAGGCTCCACTATCCCTGCCCCACCCTTAATTCTGC-
TCAGAGACCTAGAAGATCGGTGAGACAGCAGCTTGAGGCTGGCAGGGTGGTCACC-
15 CATTCCACCTTGAGCCCCACCACTGTGAGCCTCTCATTCTGACCAAGACTCGGGGAT-
TCGAACCCCTATACTACCCAAAGACTCGGCTTCTAGAGCCCCCAGTTCCAGGGGAC-
TCAGGAATTCCAGCTCCAACGTCTCCCGGGATGAAGGGGTAGAATCCCTCCATT-
CAAGAATTGAGGCATCCGAACCCGCTTTCCTTCCCTCAGTAAACAGGCAACGAGT-
TTCTTCTAAGGATCCAGGTGTGCGGCGCGCCCAAATTCCGCCCTGGGACCTGG-
20 CGTCCGAGTCCCCTCCCAATCCTCCAGGGACGCGGGTGTGGGCTTTTTCAGGGCC-
TCTGGTCCCCAGGAGGTGAAACTCACGGATCCGGGCAGATCCTGGCACCTGGGGG-
CTTCTCCAGCTCGGGCTCCGGCTTGGGAGCGGAGAACGGGCGGGCAGGAGC-
TGGGAACAGGTTAGACGACGTGACTTGGGCTGGAGGGAGGCGGGTCCCGGTGGG-
GAGGGGGAGCCAAGGTGCGCTCGAGCACCTTGGGACTTGTAGTCCCGGAGGGACAG-
25 GAGTAGCCCAAGACGATCCCATTTGGATTACCCAGAGTCCATTTACAGACAG-
GAAGGGCGAGGCCAGAGCCGAGAGCCAGGACAGGCGAGGAGATACAGAAGAGC-
CGAGACGCTGCTCGCTGTGGCTGGAGACTGACTCCTGAGCCCTTCCCCACCCCT-
TCAGGCGCACTATCCCCCTTCTGATCAGTATCCCCAGGGTCTCTGAGCCCGAATC-
TCCCCGTGATAAAAAGCGCGGGTGGATCTTCAAAGGATGTCCCAGCAAGAGTT-
30 CAAAATTTAGTTGGACTACAACCCCGAGCCCTCCGCGACCGCCCTCGGGCGAC-
TCTTTGCCTCGGGTCTGTGGGAATTGTAGTCTGGAGCCCGCAGGGCTGCACCC-
CGGTGTCTCTCTCGCCACGCGAAGGAAACCGTCTGGAGATCCTGGATAGGGGAAA-
CATTTCCCTTCCCTTGACCCTCCCTCCGCTCTGGAAGCCTCTCCACCTGGGGA-
GAAGGGTGCCCCAATTCTGGAGTAGGATCTAAATCTTGGCAGAGGGGGCGG-
35 GAAGTGGCGCTGACACACTGGCCAGGAATGCAGTGGGTCACCTGTCTAGCCAC-
CGTCTCGCGGCTCCAACCGCCGCCAACCGGGGCGGCCCCAGTGGAAGG-
GAAGTGGGTGCGTCCCCCAAATCTGTGTCCACGTGCCGCTGTTTACACGCTCCCTGG-
GGCAGGAGGAGTCCCGATCAGGTCCTTCTGAAAGTCATCGAGGTTTCCACG-
CATGAGACTAAACCCCGAGGGCATCTACAAGTCCCATTTGATCCACAACGCTACAC-
40 CGTGGCCAGCACCACTCCACGCGTGTGGGGCTCCTGGGTCCGAGGCTCCGCC-
TCGAGAACCACAAGCTCCTCCCCCTATGTTTCCCGCTCCCCCGGAGTCCAGAAGCCC-
CGCCCTGGCTGGAACCTCACGCCCTCCGGACGGATTGCCCTATTTCTCCATTTTCC-
CGCTTCTCCAGTCAAGTTCTGAAGTTGTGAGGCATCTGGGCCTCCCAGAAGACATT-
TAACACAGAAAGCACAGCCCTACTAAGTATTTCTTACCTGTCTCTTCAAGAATTTCA-
45 GACCAATCGACCGTCTGTCTTTAAGGCTTAGGAAGAGCAGTGTGGCTGCCCTT-
TAAGGAGGCGTTGCAACAAACCATATTGGACAGACGATGGGGGCGACCCATCGG-
GACCCGACGGGCTCTGACTCCAGCAATACAGCGAATCAGCGGCTTTCGGGAATA-
CATTTTTCGGAAAAAGACTTCTTCTCGGTTTTCTGCTCTGCACACGTTGAAATTTCC-
CCAGTTTTTCTGCGAGATCGGGAGTCCAGCAATGCCTACCCCGCGCTCCCGCAC-
50 CAGTTGGGCGCTCCCGGATGATGCCCTACCCCTTTGGATCCACGTGGTCTGCAACCT-
GGTGGAGCAGCCCCGGGCTACAGGGTTGCCTGAGGTGTGGGTCCAGGATGGAG-
GAGCCCCAGGCCGGCGGTGAGGGTGCGGGTTGACGGGGTGCAGAGGGTGCCTTG-
GTGGAAGGAGAAAGGGGCGTCCGAGAGGGTTCCGGCGGAAAGGAGGCGTACCTG-
CAAGCAGGACTTGCAGAGAGCGTGCATTCAGTGGGCGAACGGGAATTCGAACG-
55 GAGAGAGGGTTATCTTGTGGGGCTACCCGTGGAGAGCAAGGCGCCCCCAGGG-
GTTGGATCGGTGAAATTGAGGTGCGCCCTGGGGAACAGGTGGGCAAGAGGA-
GAAACCAGGTTGAGGGGACTGGAGTGCTCACGAGGTTAAGACCAATGGACCGA-
TAGGCGCGCCCTGCAAGATTGGACCGCAAGGAGGTGTAGTCCAGCCCATTTCCCC-

P 687 DK01

57

TTCTGCTGCAGATGCTGCTCGGTTCTCTTGTCCTCCCACTTTACCGCGAAGCCCC-
 CAGCCTCAGAGTCCCCTCGTTTCTCCTTGGAGGCGCTGACGGGTCCAGATACGGAGC-
 TGTGGCTTATTCAGGCCCTGCAGACTTTGCCCCAGAATGGTGAGTGGTCTTGT-
 5 GACGGAAAAGAGGGTCCCGGTCCAGACCCCAAGAGCGGGTCTTGAATTTGTCACAG-
 GAAAGAATTAGAGGTGAGTCACAGAGCACAGTGAAAGAAACAAGTTTATTGAAAC-
 TACTCCTTTACAGAGTAGAGTGTCTCAGAAAGCAGGGGGAGAAACCCACAGCCCT-
 TTGTTAGTATTTCTACTTATAAGAACTATAAGGAACTATAGTTAAACTTGGAGTGTG-
 CAGATAAGCTCACTAAAGGTAGGGGCTATTGGTGTATCCACGACCATTAACTCTG-
 10 CAACCTAAGCTTGCTCATTATGTTATATTTAAGTAATGGGGGCTGCATTCTTAGGA-
 CATTGGACATTCTGCAGGCTTGGTGGAAACATGTTCTGTATGGCCATAAATATTCTGT-
 ATTATAATTGGTGGTTCAGCCTGGGATGTGGTTATTTTCAGGCCATAAGCATGAACCTT-
 GTAAGTGCCTAGCTACTCACTTTAAGATGGAGTCACTCTAGTCATGTTTTATTAATAAC-
 CAGAGGCCAGCCAGGCGCAGTGGCTGGTGTGCTGTAAATCCCATCCTTTGGGAGGC-
 15 CGAGGCGAGCAGATCACTTGAGGTGAGGATTCAGGCTTCAAGACCAGCCTGGCCAACTAGT-
 GAAATTGTCTCTACTAAAAATACAAAAATTGGCTGGGCGTGGTGGCAGGTGCCTGTA-
 ATCCCAGCTACTTGAGAGGCTGAGGCAGGAGAATCGCTTGAACCCAGGAGGTGGA-
 CATTGCAGTGAGCCGAGATCATGCCACTGCACTCCAGCCTAGGCAACAGAGCAAGAC-
 TCTCTCAAAAAAATAACAAAAAATCAAAAAACCTTCCCTCTCCTGTTCCACTTAAG-
 20 CCTCTGCCCTCCCTGTTTCTCTCTGTAGCTTCAATGGGCGGCATGTGCTCTCTGG-
 CTCCCAGATCGTCAAGGGCAAATTGGCAGGCAAGCGGCACCGCTATCGAGTCTCAG-
 CAGCTGTCCCCAAGCTGGAGAAGCGACCCTGCTGGCCCCCTCAACGGAGGCAGGAG-
 GTGGACTCACCTGTGCCTCAGCCCCCAGGGCACCTAAGGATCCTTGAGGGTCCC-
 CAGCAATCCCTGTGAGGAGCCCTCTGCAGCCATCCCAGCAAGTCCCCACCACAG-
 25 GATCCCTCCTGGCCTGAGGCCTCGGTTCTGTGCCTTTGGGGCAACCCACCAGTCA-
 CAGGGCCTAGGTGAGCCTTGGCCCCCAACCTGCTCACCTCAGGGAAGAAGAAAAAG-
 GAGATGCAGGTGACAGAGGCCCTCACTCAGGAGGCAGTGAATGGGCACGGGGG-
 CCTGGAGGTGGACATGGCTTTGGGGTCGCCAGAAATGGATGTGCGGAAGAAGAA-
 GAAGAAAAAATCAGCAGCTGAAAGAACCAGAGGCAGCAGGGCCTGTGGGGACA-
 30 GAGCCACAGTGGAGACACTGGAGCCTCTGGGAGTGCTGTTCCCGTCCACCACCAA-
 GAAGAGGAAGAAGCCCAAAGGGAAAGAAACCTTCGAGCCAGAAGACAAGACAGT-
 GAAGCAGGAACAGATTAACTGAGCCTCTAGAAGACACAGTCTGTCCCCGAC-
 CAAAAAGAGAAAGAGGCAAAAGGGGACGGAAGGGATGGAGCCAGAGGAGGGGT-
 GACAGTTGAGTCTCAGCCACAGGTGAAGGTGGAGCCACTGGAGGAAGCCATCCCTCT-
 35 GCCCCCTACGAAGAAGAGGAAAAAGAAAGGGGACAGATGGCAATGATGGAGCCAG-
 GGACGGAGGCATGGAGCCAGTGGAGCCGAGATGAAGCCTCTGGAGTCCCCAGG-
 GGGGACCATGGCGCCTCAACAGCCAGAAGGAGCGAAGCCTCAGGCCACGGCAGCTC-
 TGGCAGCTCCCCAAAAGAAGACGAAGAAAGAAAAACAGCAAGATGCCACAGTGGAGC-
 CAGAGACAGAGGTGGTGGGGCCTGAGCTGCCGGATGACCTTGAGCCTCAGGCAGC-
 40 TCCCACATCCACCAAGAAGAAGAAGAAGAAAGAGAGAGGTGACACAGTGA-
 GAGCCAATTCAGCCACTAGAGCCTGAACTGCCAGGGGAGGGACAGCCTGAAGCCAG-
 GGCAACTCCGGGATCCACCAAGAAGAGGAAGAAGCAGAGTCAGGAAAGCCGGATGC-
 CAGAGACAGTGGCCCAAGAGGAGATGCCAGGGCCGCACTGAATTCAGAGTCTGGG-
 GAGGAGGCTCCACAGGCCGGGACAAGAAGCGGAAGCAGCAGCAGCAGCAGCCT-
 45 GTGTAGTCTGCCCCCGGGAAACTGAGGAACAAAGAAAGCTGAAGGTGCCACCTG-
 GGCCACCAGAAGGTGACACCCCAAGAATCCCTCCCAGAGACTGCACCAGCGCAGCC

Example 7

The cases and controls in example 6 had been individually matched with respect to
 50 age, menopausal status and hormone treatment. Therefore, it was possible to make
 a paired analysis. This generally reduces the possibility of bias and confounding, but
 often produces less significant results. When the "high-risk" group was analysed, i.e.
 RAI1^{AA} ASE-1e3^{GG} ERCC1^{GG}, versus all other genotypes, we found a rate ratio

P 887 DK01

58

(RR) = 1.64, Confidence Interval (CI) = 1.17-2.29, and with a level of significance $p = 0.004$. Thus, the "high-risk" genotype was clearly overrepresented among the breast cancers.

5 Example 8

10 In the data of example 7, the "high-risk" group was further analysed, i.e. RAI1^{AA} ASE-1e3^{gg} ERCC1^{gg}, versus all other genotypes, among those pairs that were less than 55 years of age. This increased the difference dramatically, indicating that the high-risk genotype predisposes to early breast cancer (rate ratio (RR) = 9.5, Confidence Interval (CI) = 2.21-40.79, and with a level of significance (p) = 0.003). In older age brackets, the RR was still above 1, but not significantly so. Thus, the combination of the three SNPs allows for the definition of a high-risk group for early breast cancer.

P 887 DK01

59

Claims:

1. A method for estimating the cancer risk of an individual comprising
 - 5 - providing a sample from said individual,
 - assessing in the genetic material in said sample a sequence polymorphism
 - in a region corresponding to SEQ ID NO: 1, or a part thereof, or
 - 10 - in a region complementary to SEQ ID NO: 1, or a part thereof, or
 - in a transcription product from a sequence in a region corresponding to SEQ ID NO: 1, or a part thereof, or
 - 15 - or translation product from a sequence in a region corresponding to SEQ ID NO: 1, or a part thereof,
 - obtaining a sequence polymorphism response,
 - estimating the cancer risk of said individual based on the sequence polymorphism response.
- 20 2. The method according to claim 1, wherein the cell sample is a blood sample, a tissue sample, a sample of secretion, semen, ovum, a washing of a body surface, such as a buccal swap, a clipping of a body surface, including hairs and nails.
- 25 3. The method according to any of the preceding claims, wherein the cell is selected from white blood cells and tumor tissue.
4. The method according to any of the preceding claims, wherein the sequence polymorphism comprises at least one mutation base change.
- 30 5. The method according to any of the preceding claims, wherein the sequence polymorphism comprises at least two base changes.
- 35 6. The method according to any of the preceding claims, wherein the sequence polymorphism comprises at least one single nucleotide polymorphism.

P 687 DK01

60

7. The method according to any of the preceding claims, wherein the sequence polymorphism comprises at least two single nucleotide polymorphisms.
- 5 8. The method according to any of the preceding claims, wherein the sequence polymorphism comprises at least one tandem repeat polymorphism.
9. The method according to any of the preceding claims, wherein the sequence polymorphism comprises at least two tandem repeat polymorphisms.
- 10 10. The method according to any of the preceding claims, wherein the cancer is selected from skin carcinoma including malignant melanoma, breast cancer, lung cancer, colon cancer and other cancers in the gastro-intestinal tract, prostate cancer, lymphoma, leukemia, pancreas cancer, head and neck cancer, ovary cancer and other gynecological cancers.
- 15 11. The method according to any of the preceding claims, wherein the cancer is selected from skin cancer, lung cancer, colon cancer and breast cancer.
- 20 12. The method according to any of the preceding claims, wherein the cancer is selected from skin cancer and breast cancer.
13. The method according to any of the preceding claims 10-12, wherein the skin cancer is basal cell carcinoma.
- 25 14. The method according to any of the preceding claims, wherein the assessment is conducted by means of at least one nucleic acid primer or probe, such as a primer or probe of DNA, RNA or a nucleic acid analogue such as peptide nucleic acid (PNA) or locked nucleic acid (LNA).
- 30 15. The method according to claim 14, wherein the nucleotide primer or probe is capable of hybridising to a subsequence of the region corresponding to SEQ ID NO: 1, or a part thereof, or a region complementary to SEQ ID NO:1.

P 687 DK01

61

16. The method according to claim 14, wherein the primer or probe has a length of at least 9 nucleotide or peptide monomers.

5 17. The method according to any of the preceding claims 14-16, wherein at least one primer or probe is capable of hybridising to a subsequence selected from the group of subsequences

1. GCTCTGAAAC TTACTAGCCC(A/G)GTATTTATGG AGAGGCATTT
2. GTGGTCAAAT TCTCATTCAT CGTGG (T/C) CCAGGCAAGC
10 ACACTTCCTC
3. ACCCTGAGGT GAGCACCTGT TCCTT(C/T) TCCTTGCCCT TAGCCCA-
GAG GTAGA
4. GGGCAGGGGT TTGTGCCTCC AATGA (G/A) CACAAGCTCC
CCCTGCCCCC CAACT
- 15 5. CCTGGCGGTG GCCGTCACCA GCTTT (T/C) GGGGGTGTTT
GGGAAGCTGG
6. CTCCAGCCCC ACTGTTCCCT (A/G) GGCCCTATTG GTCCCCCTGG
7. ACAAGGAGGA GGCAGAAGTG AGGTT (G/C) AAACCCACTG CCCAATC-
TTA
- 20 8. CCAACACGGT GAAACCCCGT CTGTA(T/C)TAAAAATACA AAAATTAGCC
9. AATCCAGGAC CCCATAATCT TCCGT (C/T) ATCTAAAACA ATA-
ATGGTGA
10. CCCAAGGGGG CGAGGGGAGG GTGAA (A/G)GGGTGGGACG
GGGGCAGCCG
- 25 11. GAAGTGAGAA GGGGGCTGGG GGTCG (G/-) CGCTCGCTAG
CGGGCGCGGG
12. CGCACGCGCA GTATCCCGAT TGGCT (C/G)TGCCCTAGCG GATT-
GACGGG
13. AACTCCTGGG TTCGATCAAT ACTCA (GACA/-) ATCTTGGCAG
30 GCGCAGGAGG
14. GCTGGGATTA CAGGCTTGAG CCACC (A/G) CGCCCGGCCT
GCAAAGCCAT
15. TTTTGTATCT TTAGTAGAGA CAGG (T/G) TTTCTCCATG TTGGTCAGGC
- 35 16. GCCTCAGCCT CCCGAGTAGC TGAGACT (C/A) CAGGTGCCCCG CCAC-
CACGCC

P 687 DK01

62

17. TGAAATTGTA GGTGAGAGG CCAGGCG (C/T) GGTGCTCACG
CCTGTAATTT
18. GTTTATAAAC ATTAACCAG (T/A) GCTGTGTGAA GGCACCTAAT
19. CCGTCTCTAT TAAAAATATA AAA (A/C) AATTTAGCCG GGTGTAGCGG
- 5 20. GGGAGGCTCG AGGCGGGC (A/G) GATTGCATGA GCTCAGGATT
21. TCCCAAGTTT CAGGGCCCAA (T/G) ATTCTCAAAT CACAGGATTC
22. TGCAGTGAGC TGAGATCGC (A/G) CCACTGCACT CCAGCCTGGG
23. TCTTAGGACG CATGGGGGT (T/G) GAGAGAACGG GGAGATAGAC
24. CTGGGTCTA GAACTACC (C/T) ATGCAAACCC AGCTGTTTCC
- 10 25. ATTCTGCCCT GGGTTCTAGA ACTACCT (C/A) TGCAAACCCA
GCTGTTTCCC
26. GCTGTTTCCC ACCCCATAAG GCA (A/G) TAGGGGAGCC
CACCTCCGCC
27. GACCTAGAAG ATCGGTGAG A (C/T) AGCAGCTTGA GGCTGGCAGG
- 15 28. CTGGCCAGGA ATGCAGTCGG GTCAC (C/T) CTGTCTAGCC
ACCGTCTCGC
29. GGGAGGAGTC GCCGATCAGG (C/T) CCCTTCCTGA AAGTCATCGA
30. GCAGCCCGGG CTACAGGGTT (A/G) CCTGAGGTGT GGGTCCCAGG
31. TAGAAATACT AACAAAGGGC (T/C) GTGGGTTTCT CCCCCTGCTT
- 20 32. ACAGGAGAGG GAAGGTTTTTTG (A/T) TTTTTTTTTT GTTTTTTTTT
33. GAAGAGGAAG AAGCCCAAAG GGA (A/C) AGAAACCTTC GAGCCA-
GAAG
34. GCGCCTCAAC AGCCAGAAGG AGCG (A/G) AGCCTCAGGC CCAGG-
CAGCT
- 25 35. TTGAGACTCT CTGTTTGAT (A/G) CTTCACTCAG AAGGTGCTTC
36. AGGCCAGGCT CCTGCTGGCT G (C/G) GCTGGTGCAG TCTCTGGGGA
37. CCCCTATACC CTCAAGCAT (C/T) TATCCATTGA GTTACAAACA
38. ACCATCCCCC GCCTTCCGTT (A/C) GTCCGGCCCC CGAGGCTAGC
- 30 or to a sequence complementary to any of the subsequences.

18. The method according to claim 17, wherein at least one nucleotide probe is selected from the group consisting of

P 687 DK01

63

1. TGAAATTGTA GGTGAGAGG CCAGGCG (C/T) GGTGCTCACG
CCTGTAATTT
2. GTTTATAAAC ATTAAACCAG (T/A) GCTGTGTGAA GGCACCTAAT
3. CCGTCTCTAT TAAAAATATA AAA (A/C) AATTTAGCCG GGTGTAGCGG
- 5 4. GGGAGGCTCG AGGCGGGC (A/G) GATTGCATGA GCTCAGGATT
5. TCCCAAGTTT CAGGGCCCAA (T/G) ATTCTCAAAT CACAGGATT
6. TGCAGTGAGC TGAGATCGC (A/G) CCACTGCACT CCAGCCTGGG
7. TCTTAGGACG CATGGGGGT (T/G) GAGAGAACGG GGAGATAGAC
8. CTGGGTTCTA GAACTACC (C/T) ATGCAAACCC AGCTGTTTCC
- 10 9. ATTCTGCCCT GGGTCTAGA ACTACCT (C/A) TGCAAACCCA
GCTGTTTCCC
10. GCTGTTTCCC ACCCCATAAG GCA (A/G) TAGGGGAGCC
CACCTCCGCC
11. GACCTAGAAG ATCGGTCGAG A (C/T) AGCAGCTTGA GGCTGGCAGG
- 15 12. CTGGCCAGGA ATGCAGTCGG GTCAC (C/T) CTGTCTAGCC
ACCGTCTCGC
13. GGGAGGAGTC GCCGATCAGG (C/T) CCCTTCCTGA AAGTCATCGA
14. GCAGCCCGGG CTACAGGGTT (A/G) CCTGAGGTGT GGGTCCCAGG.
15. TAGAAATACT AACAAAGGGC (T/C) GTGGGTTTCT CCCCCTGCTT
- 20 16. ACAGGAGAGG GAAGGTTTTTTG (A/T) TTTTTTTTTT GTTTTTTTTT
17. GAAGAGGAAG AAGCCCAAAG GGA (A/C) AGAAACCTTC GAGCCA-
GAAG
18. GCGCCTCAAC AGCCAGAAGG AGCG (A/G) AGCCTCAGGC CCAGG-
CAGCT

25

or to a sequence complementary to any of the subsequences.

19. The method according to claim 18, wherein at least one nucleotide probe is se-
lected from the group consisting of

30

1. GTTTATAAAC ATTAAACCAG (T/A) GCTGTGTGAA GGCACCTAAT
2. CCGTCTCTAT TAAAAATATA AAA (A/C) AATTTAGCCG GGTGTAGCGG
3. GGGAGGCTCG AGGCGGGC (A/G) GATTGCATGA GCTCAGGATT
4. TCCCAAGTTT CAGGGCCCAA (T/G) ATTCTCAAAT CACAGGATT
- 35 5. TGCAGTGAGC TGAGATCGC (A/G) CCACTGCACT CCAGCCTGGG

P 687 DK01

64

or to a sequence complementary to any of the subsequences.

5 20. The method according to any of the preceding claims, wherein at least one sequence polymorphism is assessed in a region corresponding to SEQ ID NO: 1 position 1521-37752 (t).

10 21. The method according to any of the preceding claims, wherein at least one sequence polymorphism is assessed in a region corresponding to SEQ ID NO: 1 position 7760-22885 (RAI).

22. The method according to any of the preceding claims, wherein at least one sequence polymorphism is assessed in a region corresponding to SEQ ID NO: 1 position 34391- 37752.

15 23. The method according to any of the preceding claims, wherein at least two different probes are used, one probe being selected from the probes as defined in any of claims 17-21, and the other probe being capable of hybridising to a sequence different from SEQ ID NO: 1, or a part thereof, or to a sequence complementary to a region different from SEQ ID NO: 1, or a part thereof.

20 24. The method according to claim 1, wherein the translational product from a sequence in a region corresponding to SEQ ID NO: 1, or a part thereof, is an antibody, such as a monoclonal or polyclonal antibody.

25 25. A method for estimating the cancer prognosis of an individual comprising

- providing a sample from said individual,
 - assessing in the genetic material in said sample a sequence polymorphism
- 30
- in a region corresponding to SEQ ID NO: 1, or a part thereof, or
 - in a region complementary to SEQ ID NO: 1, or a part thereof, or
 - in a transcription product from a sequence in a region corresponding to SEQ ID NO: 1, or a part thereof, or

P 687 DK01

65

- or translation product from a sequence in a region corresponding to SEQ ID NO: 1, or a part thereof,
 - obtaining a sequence polymorphism response,
- 5 - estimating the cancer prognosis of said individual based on the sequence polymorphism response.

26. The method according to claim 25, wherein the method has any of the features as defined in any of the claims 2-24.

10

27. A method for estimating a treatment response of an individual suffering from cancer to a cancer treatment, comprising

15

- providing a sample from said individual,
 - assessing in the genetic material in said sample a sequence polymorphism
- 20 - in a region corresponding to SEQ ID NO: 1, or a part thereof, or
- in a region complementary to SEQ ID NO: 1, or a part thereof, or
 - in a transcription product from a sequence in a region corresponding to SEQ ID NO: 1, or a part thereof, or
 - or translation product from a sequence in a region corresponding to SEQ ID NO: 1, or a part thereof,
 - obtaining a sequence polymorphism response,
- 25 - estimating the individual's response to the cancer treatment based on the sequence polymorphism response.

30

28. The method according to claim 27, wherein the method has any of the features as defined in any of the claims 2-24.

29. A primer or probe for use in a method as defined in any of the claims above, said primer or probe being selected from

35

TGGCTAACACGGTGAAACC(SEQ ID NO:7)

P 687 DK01

66

5 GGAATCCAAAGATTCTATGATGG (SEQ ID NO:8)
GGGAGGCGGAGCTTGCACTGA (SEQ ID NO:9)
CTGAGATCGCACCCTGAC (SEQ ID NO:10)
GGTTTTCTGCTCTGCACACG (SEQ ID NO:11)
CCTTTCTCCTTCCACCAACG (SEQ ID NO:12)
CGGGCTACAGGGTTACCTGAG (SEQ ID NO:13)
TCTGCAACCTGGTGCAGCAGC (SEQ ID NO:14)
CCTACCACCATCATCATCC (SEQ ID NO:15)
GCCTTGCCAAAATCATAACC (SEQ ID NO:16)
10 CCTCTCCCAATTAAGTGCCTTCACACAGC (SEQ ID NO:17)
AGCCAGGGAGGTTGAGGCT (SEQ ID NO:18)
AGACAGCCCTGAATCAGCAC (SEQ ID NO:19)
GCAATGAGCCGAGATAGAA (SEQ ID NO:20)
TGGCTAGCCCATTACTCTA (SEQ ID NO:21)

15

30. A primer or probe for use in a method as defined in any of the claims above as the other probe

20 GCCCCGTCCCAGGTA (SEQ ID NO:21)
AGCCCCAAGACCCTTCACT (SEQ ID NO:22)
GTCCCATAGATAGGAGTGAAAG (SEQ ID NO:23)
CCCTAGGACACAGGAGCACA (SEQ ID NO:24)
TTGTGCTTTCTCTGTGTCCA (SEQ ID NO:25)
TATCAGAAAAGGCTGGAGGA (SEQ ID NO:26)
25 GAGTGGCTGGGGAGTAGGA (SEQ ID NO:27)
GCCAAGCAGAAGAGACAAA (SEQ ID NO:28)
CCTCAGATGTCCTCTGCTCA (SEQ ID NO:29)
GCCACAGCCCCAGCAAGTAG (SEQ ID NO:30)
AGGACCACAGGACACGCAGA (SEQ ID NO:31)
30 CATAGAACAGTCCAGAACAC (SEQ ID NO:32)
TTAGCTTGGCACGGCTGTCCAAGGA (SEQ ID NO:33)
ACAGAATTCGCCCCGGCCTGGTACAC (SEQ ID NO:34)
TTGAACTGGAACTCTGAGAAGG (SEQ ID NO:35)
TGGTGGATGGTGTGAAGCA (SEQ ID NO:36)
35 CCTTTCTCCAATTCTTCTCCATTTCACC (SEQ ID NO:37)

P 687 DK01

67

GGGGATCATGTCGTCAATGGACT (SEQ ID NO:38)
ATGCCCTGTAGGTTCAATGG (SEQ ID NO:39)
TGGAGGTCTTTAGGGGCTTG (SEQ ID NO:40)
GGCTGGTCCCCGTCTTCTCCTTCC (SEQ ID NO:41)
5 TCTCTGTTGCCACTTCAGCCTC (SEQ ID NO:42)
GTCCTGCCCTCAGCAAAGAGAA (SEQ ID NO:43)
TTCTCCTGCGATTAAAGGCTGT (SEQ ID NO:44)
ATCCTGTCCCTACTGGCCATTC (SEQ ID NO:45)
TGTGGACGTGACAGTGAGAAAT (SEQ ID NO:46)
10 TGGAGTGCTATGGCACGATCTCT (SEQ ID NO:47)
CCATGGGCATCAAATTCCTGGGA (SEQ ID NO:48)
CACACCTGGCTCATTTTTGTAT (SEQ ID NO:49)
TCATCCAGGTTGTAGATGCCA (SEQ ID NO:50)
AGGCTCAACAAGGAAAAATGC (SEQ ID NO:51)
15 GCTAGACAGTCAAGGAGGGACG (SEQ ID NO:52)
AAAGGGTGGGTGTGGGAGACATTGG (SEQ ID NO:53)
AAACCAACCTAGGCACCCCAA (SEQ ID NO:54)
CAGTGTCCAAAGAGCACC (SEQ ID NO:55)
CTACCCCTTTAGCGACC (SEQ ID NO:56)
20 TCCTGCCCCCAGAGCGTCACC (SEQ ID NO:57)
GTACGGTCCACATAATTTTGGAGGA (SEQ ID NO:58)
CGACGAACTTCTCTGAAGCGAA (SEQ ID NO:59)
AGCGACACGGGCATCTGG (SEQ ID NO:60)
ATGAGCGTCCACCTCCTGAACC (SEQ ID NO:61)
25 AGGCAGCAGCATCGTCATCCCC (SEQ ID NO:62)
TGCATAGCTAGGTCCTGC (SEQ ID NO:63)
AACTGACRAAACTAGCTCTATGGGGTGGTGCCGCA (SEQ ID NO:64)
CTGGCTCTGAACTTACTAGCCC (SEQ ID NO:65)
GCTGGACTGTCACCGCATG (SEQ ID NO:66)
30 GGAGCAGGGTTGGCGTG (SEQ ID NO:67)
TGCCCTCCCAGAGGTAAGGCCT (SEQ ID NO:68)
CCCTCCCGGAGGTAAGGCCTC (SEQ ID NO:69)
GATCAAAGAGACAGACGAGC (SEQ ID NO:70)
GAAGCCCAGGAAATGC (SEQ ID NO:71)
35 GGACGCCCACCTGGCCAACC (SEQ ID NO:72)

P 687 DK01

68

CGTGCTGCCCAACGAAGTG (SEQ ID NO:73)

- 5 31. The primer or probe according to any of claims 28 or 29, wherein the probe is operably linked to at least one label, such as operably linked to two different labels.
- 10 32. The probe according to claim 30, wherein the label is selected from TEX, TET, TAM, ROX, R6G, ORG, HEX, FLU, FAM, DABSYL, Cy7, Cy5, Cy3, BOFL, BOF, BO-X, BO-TRX, BO-TMR, JOE, 6JOE, VIC, 6FAM; LCRed640, LCRed705, TAMRA, Biotin, Digoxigenin, DuO-family, Daq-family.
- 15 33. The primer or probe according to any of claims 28-31, wherein the primer or probe is operably linked to a surface.
34. The primer or probe according to claim 32, wherein the surface is the surface of microbeads or a DNA chip.
35. An antibody directed to an epitope of a RAI gene product.
- 20 36. A kit for use in a method as defined in any of the claims above, comprising at least one primer or probe, said probe being as defined in any of claims 29-35, and optionally further amplifying means for nucleic acid amplification.

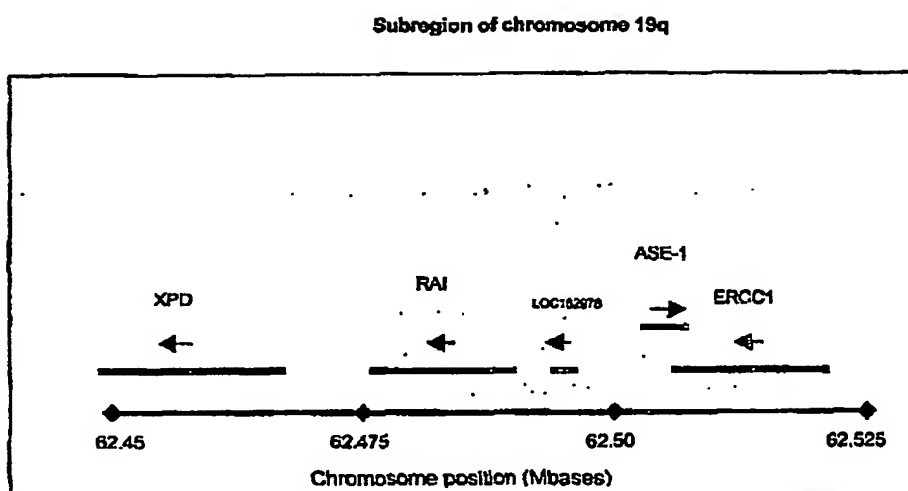
1/3

Modtaget

7 OKT. 2002

PVS

Fig. 1



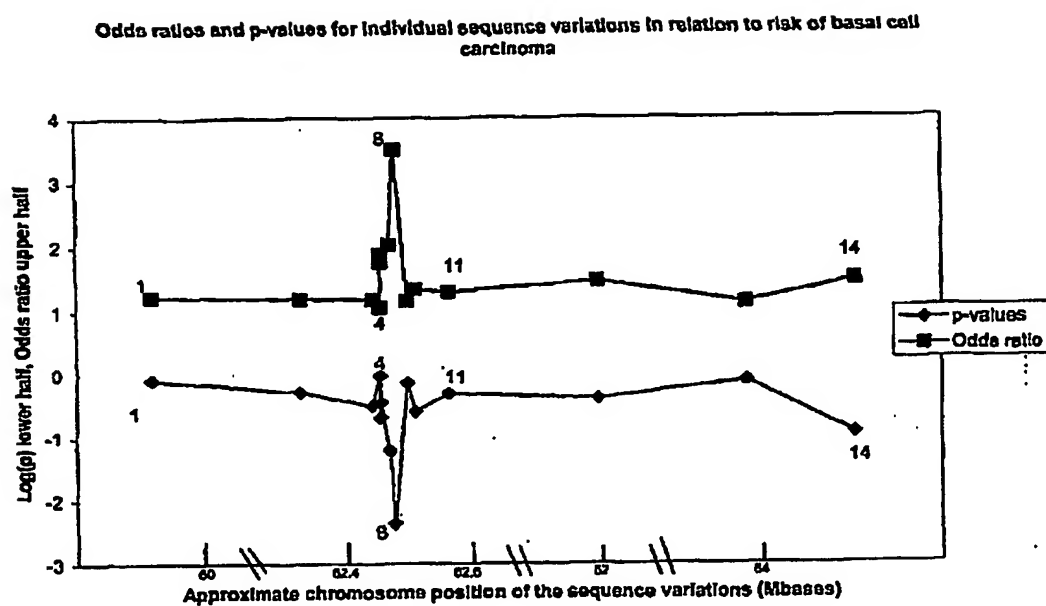
2/3

Modtaget

- 7 OKT. 2002

PVS

Fig. 2



3/3

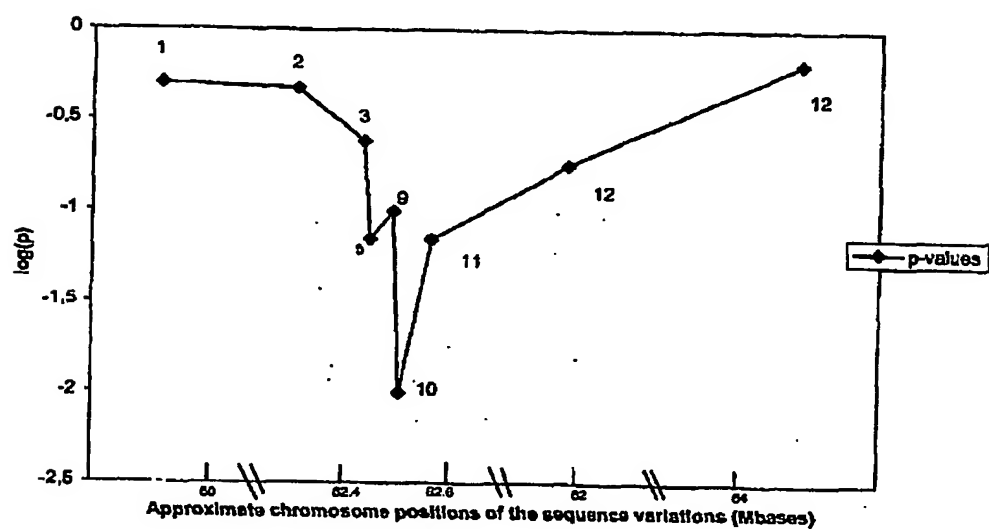
Modtaget

- 7 OKT. 2002

PVS

Fig. 3

p-values for association of different sequence variations with risk of basal cell carcinoma among psoriatic Danes



**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

☐ **BLACK BORDERS**

☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**

☐ **FADED TEXT OR DRAWING**

☒ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**

☐ **SKEWED/SLANTED IMAGES**

☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**

☐ **GRAY SCALE DOCUMENTS**

☒ **LINES OR MARKS ON ORIGINAL DOCUMENT**

☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**

☐ **OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.